



Published in final edited form as:

Nat Methods. 2013 September ; 10(9): 877–879. doi:10.1038/nmeth.2560.

A database of *C. elegans* behavioral phenotypes

Eviatar Yemini, Tadas Jucikas, Laura J. Grundy, André E.X. Brown, and William R. Schafer
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom

André E.X. Brown: abrown@mrc-lmb.cam.ac.uk; William R. Schafer: wschafer@mrc-lmb.cam.ac.uk

Abstract

Using low-cost automated tracking microscopes, we have generated a behavioral database for 305 *C. elegans* strains, including 76 mutants with no previously described phenotype. The database consists of 9,203 short videos segmented to extract behavior and morphology features that are available online for further analysis. The database also includes summary statistics for 702 measures with statistical comparisons to wild-type controls so that phenotypes can be identified and understood by users.

The principal output of the nervous system is motor behavior. Therefore, a variety of neural perturbations ultimately manifest as changes in motion, making locomotion a useful phenotype for neurogenetics in model organisms. For the nematode *C. elegans*, the study of mutations that cause visible defects in spontaneous crawling have given insight into neural functions as diverse as synaptic transmission, axon guidance, neuromodulation, and proprioception. Despite these successes, extending behavioral phenotyping to large-scale screens remains a challenge. In fact, 85% of *C. elegans* genes have no reported phenotypic effect when knocked down using RNA interference, though most genes are required for wild-type fitness¹. These findings suggest that there is a *phenotyping gap*: knocking down most genes has an effect, but one too subtle to see by manual observation.

To close this phenotyping gap, it is desirable to have a system that is capable of measuring phenotypes both extensively and intensively². For behavior, extensive phenotyping requires measuring many parameters that quantify motion, posture, and path as well as the frequencies and intervals between relevant behaviors such as reversals and sharp turns. Achieving intensive sampling requires following individual worms with high temporal resolution over extended intervals. These conditions can be met using single-worm trackers that follow freely-behaving worms, using a motorized stage to keep the worm within the camera's field of view³.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Author Contributions

E.Y. created the WT2 system (hardware and software), analyzed data, and wrote the paper. T.J. created WT2 analysis pipeline and database software. L.J.G. conducted all experiments. A.E.X.B. analyzed data, supervised research and wrote the paper. W.R.S. designed experiments, supervised research and wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

Here we report extensive and intensive behavioral phenotypes for 9,203 individuals, representing 305 strains of *C. elegans* (Supplementary Table 1). Included in this dataset are most canonical *unc* and *egl* mutants as well as representative knockouts from receptor, channel and neuropeptide gene families with likely roles in nervous system function. The database contains 76 mutants with no previously characterized phenotype, 15 genes with multiple alleles, and 13 double or triple mutant combinations (the majority of which are accompanied by single mutant representation as well). The videos are easily accessible online at the video-sharing website YouTube. The data are also available online with various degrees of processing, from the skeleton and outline coordinates to the time series of extracted features, their histograms, and an in-depth view of their summary statistics. For computational researchers, the database is a rich source of processed measures and raw data to develop new algorithms for segmentation, behavioral quantification, and bioinformatic approaches which link complex phenotypes with genetic perturbations. For neurogeneticists, the summary statistics and visualizations make it possible to identify behavioral phenotypes in mutants of interest.

We collected data using Worm Tracker 2.0 (WT2), which employs a mobile camera to automatically track and record single worms (Fig. 1a). Worms were tracked on food, on an immobile platform isolated from tracking motion. Videos were analyzed to extract the worm's contour and skeleton (Fig. 1b). The head was identified automatically from the image, while the dorsal-ventral orientation was obtained from user input. Contour and skeleton data was further reduced by extracting relevant features based on common behavioral assays⁴ in addition to automated methods developed for use in previously-described worm trackers⁵⁻⁹. We developed several new algorithms for feature extraction to quantify parameters related to motion state, crawling, foraging, dwelling, and exploratory range. At the top level, the extracted features include measures of morphology, posture, locomotion, and path dynamics (see Supplementary Fig. 1 for illustrations, Supplementary Fig. 2 for an example strain comparison, and Supplementary Table 2 for the full list of measurements). These primary features were also evaluated in different contexts to give more complex parameterization; for example, mean speed was measured over the entire movie, as well as independently for periods when the animal was moving either forward or backward. Likewise, dorsal and ventral bending were measured over the entire body, and in specific regions such as the head, tail, and midbody. Finally, specific behavioral events such as reversals or omega turns were used to generate secondary parameters; for example, the frequency, time spent in execution, and the distance covered during the event. Altogether, the permutation of these measures yielded 702 distinct feature measurements. By running eight units in tandem, we could achieve a throughput of approximately 128 recordings (8 recordings each of 15 strains plus wild-type control) per day.

The data from these experiments have been compiled, and are available publicly at <http://wormbehavior.mrc-lmb.cam.ac.uk>. The data from each individual is available as a time series of features. The WT2 feature viewer provides a visual assessment of these feature time series, frame-by-frame, accompanied by the worm's contour, skeleton, and its location on plate (Fig. 1b-e). Each video is also provided, via YouTube, with an overlay of the skeleton, contour, head, ventral side, and annotations for unsegmented frames (Fig. 1b).

These two distinct views provide complementary methods to verify the fidelity of the features. Furthermore, aggregate measurements and statistics per strain are available directly on the website, presented graphically in .PDF files, in .MAT files, and as spreadsheet compatible .CSV files (see Methods). Inclusion of the skeleton in the .MAT files permits new features to be rapidly computed from the existing data set.

We compared each strain in the database to the lab-stock wild type N2 strain. Because of month-to-month variability in the N2 data (Supplementary Figure 3), each mutant strain was compared to an N2 control data set of videos collected within a two-week window centered around the date the mutant data were collected. When a measure was detected exclusively in either the strain or its control (for example, some strains never perform reversals), we used Fisher's exact test to assess significance. For the remaining cases, we used the Wilcoxon rank-sum test, controlling the false-discovery rate with a q-value in place of p^{10} . All strains were significantly different from wild type at $q = 0.05$, including those representing 76 genes with no previously characterized phenotype (Supplementary Tables 3–5).

The feature measurements in the database appear to provide a useful description for a wide range of behavioral phenotypes (Supplementary Fig. 4). For visibly uncoordinated mutants (Figure 2A), the measured locomotion features serve as a sensitive fingerprint for phenotypic similarity. Loss-of-function alleles of the same gene (e.g. *unc-2*, *egl-21*, *unc-4*, *unc-108*, *unc-89*) exhibit highly similar phenotypic signatures, as do mutants affecting genes encoding components of known molecular complexes (e.g. *unc-7/unc-9*¹¹, *unc-79/unc-80*¹², *unc-38/unc-63*¹³). It is reasonable to suppose that the signatures of uncharacterized mutants (such as *egl-31* and *egl-33*) might provide insight into the molecular or cellular functions of the mutant gene. More detailed clustering results are in Supplementary Figure 5.

With regard to phenotypes identified in previously uncharacterized knockout strains (Figure 2B), the paucity of multiple knockout alleles makes attribution of phenotypes to many gene deletions provisional. Nonetheless, some genes are represented by multiple alleles, including the previously uncharacterized *trpa-2*, a member of the TRPA family of cation channels that function in nociception and thermosensation in many animals including *C. elegans*¹⁴. TRPA-2 is an unusual member of the TRPA family as it contains no predicted ankyrin repeats. From our analysis of three *trpa-2* deletion mutants (*ok3189*, *tm3085*, and *tm3092*), we determined that all three showed a similar significantly altered posture during reversals (Supplementary Figure 6). These results suggest a potential proprioceptive role for *trpa-2*, a hypothesis that can be investigated further. We also detected new locomotion phenotypes for multiple alleles of several genes previously implicated in other processes (Supplementary Figure 6), including the TRPV channel OCR-4¹⁵ and the TRPC channel TRP-2¹⁶. These results show that automated tracking can detect previously unnoticed locomotion phenotypes, even in relatively well-studied strains.

These resources provide an entry point for understanding the functions of a substantial number of genes that have been mutated by the *C. elegans* knockout consortia but not yet phenotypically characterized. Detailed feature measurements provide information about the behavioral processes affected by the gene of interest, which may suggest a neural basis for observed phenotypes. Finally, the availability of raw video data will make it possible for

investigators to develop their own approaches to studying the behavior of particular mutant strains.

Beyond adding more strains to the phenotype database, we also hope to increase the richness of the data available for each genotype. One approach would be to record each strain under a variety of different environmental conditions, probing different sensory modalities and behavioral responses. A recently published study of behavioral responses to thermal stimuli highlights the potential effectiveness of this approach¹⁷. In addition, it should be possible to extract additional features from the existing video data. This could involve developing new algorithms for feature measurement, or more unsupervised approaches based on time series motifs¹⁸. It is our hope that making these data available to the community will encourage other researchers to contribute collaboratively to the emerging science of behavioral informatics.

Methods

Collection of worm video data

Worms were maintained under strictly controlled conditions up until the point of tracking¹⁹. To ensure sufficient phenotypic sampling, we filmed at least 20 young-adult hermaphrodites per strain for 15 minutes, spontaneously behaving on food. The camera magnification was set between 3.5–4.5 microns/pixel (a corresponding FOV of, approximately, 2.5×2mm at 640×480 resolution) and the frame rate to 20–30 frames/second to ensure a high-resolution analysis. We observed a 30-minute wait, prior to tracking, to allow worms to habituate after being picked and moved to their tracking plate. A list of all recorded strains is provided in Supplementary Table 2.

To avoid potential room conditions that may bias measurement, recording was randomized, as best possible, across multiple trackers. Strains were matched to all N2s recorded within 1 week of their own recording date. To ensure sufficient sampling while limiting the bias against extreme coiler mutants (which suffer a large portion of segmentation failures), videos with less than 3 minutes worth of segmented frames were ignored when measuring collective statistics.

Dorsal-ventral annotation and head-tail detection

Worm features necessitate dorsal-ventral and head-tail distinctions. We annotated the ventral side for each video by eye. We did not profile rolling mutants and therefore expected worms to maintain their dorsal-ventral orientation. Nevertheless, 126 random videos were examined and the worms therein found never to flip sides. Head-tail orientation was annotated automatically by software. We examined 133 random videos (roughly 1% of our data, 2.25 million segmented frames), representing a broad range of mutants (including several nearly motionless UNCs). Many of these include early videos which suffered multiple dropped frames and poor imaging conditions that were later improved. We found that the head was correctly labeled with a mean and standard deviation of $94.39 \pm 17.54\%$ across individual videos and 95.6% of the video frames collectively.

Testing strain and group significance

We used Wilcoxon rank-sum to test the differences between each group's features and the wild type (329 groups of strains and time-based N2 groups by 702 measures), and Fisher's exact test for measurements found exclusively in the experiment or control groups. Since all features were measured using automated algorithms scoring was effectively blind. False-discovery rate was controlled across these comparisons, converting p-values to their q equivalents¹⁰. Each group was assigned its minimum q-value as a measure of group significance. Under this paradigm, every group was found to be significantly different from the wild type at $q = 0.05$. To ensure this was not an artifact of our methodology we chose a second method to verify our result. Since sparse sampling led to a difficulty in assessing normality, and since dimensionality outnumbered observations for all but the lab N2 data set, we measured the Hotelling T^2 statistic using a shrinkage estimation of the covariance matrix and a permutation test to determine p-values²⁰. Four strains had exclusive measures that would be penalized as a result of this method due to the absent values. For these groups we used the minimum p-value associated with their Fisher's exact test in place of computing the MANOVA. For all other groups, we ran 10,001 permutations. The p-values were corrected, controlling the false-discovery rate, to their q equivalents. Once again, every group was found to be significantly different from the wild type at $q = 0.05$, bearing out the results of our primary methodology (Supplementary Table 3).

Reproducibility and sensitivity of measurements

We performed a number of tests to assess the sensitivity and reproducibility of the measurements in our database. To assess variability within an individual animal, we tracked 25 young-adult wild-type hermaphrodites for 2 hours without the usual habituation period we perform for our regular assays. Consistent with previous reports^{16,21}, the worm speed was well-fit by an exponential decay, with a time constant $\tau = 19$ minutes ($R^2 = 0.96$). Since speed might be expected to correlate with both crawling amplitude and frequency, we investigated whether these two measures might show similar habituation. In fact, crawling frequency (titled "Absolute Midbody Frequency" in our measurements) was best fit with an exponential decay of $\tau = 19$ minutes ($R^2 = 0.71$; Supplementary Fig. 2b); whereas amplitude (titled "Absolute Midbody Amplitude" in our measurements) decayed with a longer time constant of $\tau = 48$ minutes ($R^2 = 0.59$; Supplementary Fig. 2a). Thus, the time necessary for locomotion features to adapt to steady state appears to differ, even among features affecting similar aspects of behavior.

We next assessed the between-animal variability of worms from the same wild-type stock. Individuals of a given stock, after generations of self-fertilization, would be expected to be nearly genetically identical; however, small differences in age as well as difficult-to-control environmental factors might be expected to affect behavioral data. We recorded a total of 1,218 N2 young-adult hermaphrodites over the course of three years, from 2009 to 2012. Animals were tracked from January to December, from Tuesday through Friday, and from 9am to 5pm. We chose six representative measures to analyze variability: length, forward speed (titled "Positive Midbody Speed" in our measurements), foraging amplitude (titled "Absolute Foraging Amplitude" in our measurements), reversals (titled "Backward Motion Frequency" in our measurements), coiling frequency, and exploratory range (Supplementary

Fig. 2d and 3). To assess the influence of these factors (hour, day, and month), one-way ANOVAs were performed with Bonferroni correction for 18 tests and checked for significance at $p = 0.05$. The tested groups show a mixture of normality and non-normality when using a Shapiro-Wilk test with correction for multiple comparisons. Therefore, to avoid the assumption of normality, Kruskal-Wallis tests were also performed with the same correction and α , resulting in identical significance. Since all tracked animals were identified as fourth-stage larvae the night before, we reasoned that animals tracked later in the day should be slightly older. Indeed, there was a small but significant difference in length (and other features) between animals tracked throughout the day. We measured changes in length and midbody width, over 2 hours, among the previously mentioned group of 25 young-adult, lab-stock N2s and found them well-fit with a 1% linear growth per hour (Supplementary Fig. 3), $R^2 = 0.92$ and 0.76 respectively. This may explain the observed hourly differences between animals. Likewise, we reasoned that although we control our lab to maintain 22°C , the temperature may still vary slightly with the season whereas the day of the week should have no consistent trend. Indeed, the day of the week had no significant effect on any of the tested parameters, whereas the month of the year had a slight but significant effect on all parameters but coiling (Supplementary Fig. 3). To account for seasonal effects, all data were controlled by wild-types collected within a three-week window centered around the experiments.

With the variability in mind, we computed the number of worms required to achieve statistical power in discriminating phenotypic differences. We bootstrapped Wilcoxon rank-sum tests, comparing multiple group sizes. The tested groups were chosen from our three-year collection of wild-type lab stock and compared to ones chosen from an identical collection, mean-shifted by the discriminatory amount. The findings indicate that 10 worms provide well over 90% power to discriminate two standard deviations of mean difference. Moreover, 20 worms discriminate a single standard deviation at over 80% power (Supplementary Fig. 3). For example, when testing forward speed, 20 worms discriminate a mean difference of at least $53\mu\text{m/s}$ with 90% power (assuming the compared groups share an identically-shaped distribution to our lab-stock N2).

Assessing variation between different wild-type stocks

To estimate genetically-derived variability between different lab stocks of the same strain, we compared the behavior of our lab N2 stock to one obtained from the Caenorhabditis Genetics Center. We chose four common measures (Supplementary Fig. 2d) to assess potential differences: length, speed (titled “Midbody Speed” in our measurements), foraging amplitude, and range. Speed and foraging were further subdivided into forward-backward and dorsal-ventral statistics, yielding a total of six statistical comparisons. Shapiro-Wilk tests, with $\alpha = 0.05$, indicated a mixture of normal and non-normal distributions for our 21 CGC and 27 lab N2 measures. Therefore, we compared both groups using Wilcoxon rank-sum tests with Bonferroni correction. Of all these comparisons, only dorsal-oriented foraging ($p = 6 \times 10^{-3}$) showed a significant difference between the two wild-type stocks although visually the histograms for other features show some apparent differences and further testing, using subdivided feature measurements, uncovers related significant

measures (available online in our database). Nonetheless, many features appear relatively consistent between the two wild-type stocks despite likely genetic divergence.

We also analyzed LSJ1, a more distant relative of N2. This strain is descended from the same wild isolate as N2, but the two strains have diverged following years of laboratory cultivation^{22,23}. We compared 43 LSJ1 with 107 of our lab N2 for the same six features described above. As expected, the LSJ1 showed significantly higher speed (backward $p = 2 \times 10^{-14}$; forward $p = 6 \times 10^{-14}$) and a significantly larger range of exploration ($p = 2 \times 10^{-12}$), correlating with the bordering phenotype of LSJ1. Additionally, LSJ1 was slightly (0.1 mm) but significantly shorter ($p = 2 \times 10^{-13}$) than N2. In comparison, the summary statistics for foraging are nearly overlapping ($p = 1$ for both sides) with a dorsal mean and SEM of $37 \pm 0.5^\circ$ and $36.9 \pm 0.4^\circ$, for the LSJ1 and lab N2 respectively, and ventral values of $36.5 \pm 0.3^\circ$ for both strains.

Clustering

Each strain's measures were collapsed to the mean of their observations and normalized to their wild-type controls (by subtracting the control mean and scaling by its variance). In four strains, measures were exclusively found either in the tested strain or its control. When a measure was always observed in one set but not the other, its normalization (and consequent z-score calculation) could not be computed. We reasoned that this was an extreme case that should be reflected in the representative z-score. Therefore, when a measurement was present in a strain but not its control, its z-score was imputed to be double the population maximum. Conversely, when a measurement was present in the control but excluded from the strain, the z-score was imputed to be double the population minimum (the minimum z-score was always negative).

Computer code

Unless otherwise noted, the image processing, statistics, and bioinformatics functions were available through common Java libraries (version 1.6), standard Matlab toolboxes (version 2010a), standard R packages (version 2.15), or created as needed. Several publicly available Mathworks packages significantly facilitated the analysis and figures within this paper: the videoIO toolbox by Gerald Dalley, swtest function by Ahmed Ben Saïda, fexact function by Michael Boedigheimer, export_fig function by Oliver Woodford, notBoxPlot function by Rob Campbell, and rdir function by Gus Brown.

Wormtracker 2 details and algorithms

The Worm Tracker 2.0 (WT2) hardware guide and free software are available at: <http://www.mrc-lmb.cam.ac.uk/wormtracker/>

In summary, a camera, illumination, and motorized stage are combined to follow a single worm, navigating a thin bacterial food lawn, on an agar petri dish. The software uses a closed loop wherein live video is used to guide the stage and keep the worm centered within the camera's field of view (FOV). The petri dish housing the worm is immobile, while the camera, stage, and illumination move as one. Therefore, the worm is isolated from external forces such as stage movement.

The camera is a DinoLite AM413T with zoom magnification. We use 640×480 resolution, 30fps, with a magnification that results in 3.5–4.5 microns/pixel and an FOV of roughly 2.5×2mm at a focal distance of nearly 1cm. Older videos have a frame rate of 20–30fps. New videos maintain 30fps. The illumination is a red, Philips Lumileds, side-emitting, Luxeon III Star (LXHL-FD3C). An opal diffuser provides roughly uniform lighting over the FOV. The wavelength is 627nm to avoid exciting a short-wavelength avoidance response through the *C. elegans* LITE-1 receptor. Finally, the motorized stage uses Zaber T-NA0850 linear actuators. The stage travels 5cm at up to 8mm/s in orthogonal x and y axes with a resolution of just under 0.05µm.

Worms were maintained as previously described in “Preparation of samples for single-worm tracking”¹⁹. The protocol is also available online: <http://www.mrc-lmb.cam.ac.uk/wormtracker/webcontent/trackingProtocol.pdf>

Briefly, worms are maintained at room temperature, approximately 22°C. All plates are fresh, having been poured within 1 week of use and kept at 4°C until roughly 24 hours before use. Strains are maintained on standard NGM plates²⁴ seeded with 3 drops of OP50. 6 adult worms are transferred to a new plate to maintain stocks. At least two generations are passaged, in these conditions, prior to tracking. The evening before tracking, at roughly 5pm, L4 hermaphrodites are picked to a fresh plate, 10 worms per plate. The next morning, 3.5cm low-peptone NGM plates are seeded with 20µl of OP50, in the center of the plate, and allowed to dry. This 20µl drop of OP50 is nearly circular and roughly 8mm in diameter. The L4 worms from the evening before, now young adults, are then transferred to the center of the food on the 3.5cm plates, 1 worm to a plate. The worms are given 30 minutes to habituate, and are then tracked for 15 minutes. A wild-type N2 is always tracked, at the same time, on one of the eight nearby trackers, serving as a control. Strains are tracked between 8am and 6pm, across several hours and days, and randomly assigned to any of eight available trackers. In total (prior to filtering the videos for quality) approximately 25 worms are tracked, per experimental strain, controlled by roughly 65 N2s tracked within one week of the experiments.

Worm Segmentation

Video frames are extracted using the Matlab videoIO toolbox by Gerald Dalley. There is a sharp contrast between the worm and background in our video images. Worm pixels are segmented from the background using the Otsu method²⁵ to find a threshold. The largest 8-connected component in the thresholded image is assumed to be the worm. Frames in which the worm touches the image boundaries, is too small, lacks a clear head and tail, or has unrealistic body proportions are not analyzed further. Frames containing stage movement are also removed to eliminate bad segmentations wherein the worm image may be blurred (see the section titled “Absolute Coordinates”). Given our desire for accurate and precise measures as well as the large data volume (due to a high video frame rate), we err on the side of caution and attempt to reject ambiguous segmentations rather than include them.

Once the worm has been thresholded, its contour is extracted by tracing the worm’s perimeter. The head and tail are located as sharp, convex angles on either side of the contour. The skeleton is extracted by tracing the midline of the contour from head to tail.

During this process, widths and angles are measured at each skeleton point to be used later for feature computation. At each skeleton point, the width is measured as the distance between opposing contour points that determine the skeleton midline. Similarly, each skeleton point serves as a vertex to a bend and is assigned the supplementary angle to this bend (Supplementary Fig. 4b). The supplementary angle can also be expressed as the difference in tangent angles at the skeleton point. This angle provides an intuitive measurement. Straight, unbent worms have an angle of 0° . Right angles are 90° . And the largest angle theoretically possible, a worm bending back on itself, would measure 180° . The angle is signed to provide the bend's dorsal-ventral orientation. When the worm has its ventral side concave within the bend, the bending angle is signed negatively.

Pixel count is a poor measure of skeleton and contour lengths. For this reason, we use chain-code lengths²⁶. Each laterally-connected pixel is counted as 1. Each diagonally-connected pixel is counted as $\sqrt{2}$. The supplementary angle is determined, per skeleton point, using edges $1/12$ the skeleton's chain-code length, in opposing directions, along the skeleton. When insufficient skeleton points are present, the angle remains undefined (i.e., the first and last $1/12$ of the skeleton have no bending angle defined). $1/12$ of the skeleton (Supplementary Fig. 4a illustrates division of the worm body into parts) has been shown to effectively measure worm bending in previous trackers and likely reflects constraints of the bodywall muscles, their innervation, and cuticular rigidity⁶.

Ventral Side Annotation and Head Detection

Our worm features necessitate dorsal-ventral and head-tail distinctions. The worm's ventral side was annotated for each video by eye. We did not profile rolling mutants and therefore expected worms to maintain their dorsal-ventral orientation. Nevertheless, 126 random videos were examined and the worms therein found never to flip sides. Head-tail orientation was annotated automatically by software. We examined 133 random videos (roughly 1% of our data, 2.25 million segmented frames), a collection of 100 from a quality-filtered set and 33 rejected by this filter (see the section titled "Feature File Overview"), representing a broad range of mutants (including several nearly motionless UNC's). Many of these include early videos which suffered multiple dropped frames and poor imaging conditions that were later improved. We found that the head was correctly labeled with a mean and standard deviation of $94.39 \pm 17.54\%$ across individual videos and 95.6% of the video frames collectively.

Before assigning the head and tail, videos are split into chunks in which worm skeletons can be confidently oriented with respect to each other. Chunk boundaries are set whenever there is a gap in skeletonized frames of 0.25 seconds or more. During these gaps, worm motion could make skeleton orientation unreliable. The skeletons within each chunk are aligned by determining which of the two possible head-tail orientations minimizes the distance between corresponding skeleton points in subsequent frames. When possible, we unify chunks and heal up to $1/2$ second interruptions by determining whether the worm was bent enough to achieve an omega turn and flip its orientation. If so, we trace the worm's path through its large bend to determine the new orientation. If the path cannot be confidently traced, we avoid healing and maintain separate chunks.

The head is detected in each chunk of oriented frames. The head and neck perform more lateral motion (e.g., foraging) even in uncoordinated mutants. Therefore, we measure lateral motion at both worm endpoints, across each chunk – unless the chunk is shorter than 1/6 of a second which is too short to reliably measure such motion. In our setup, the head is lighter than the tail. Therefore, we also measure the grayscale intensity at both worm endpoints, across each chunk. Linear discriminant analysis (LDA) was used on a combination of lateral motion and intensity at the worm endpoints for a training set of 68 randomly-chosen videos. This classifier was then used for the entire data set to automatically detect and label the worm's head.

Absolute Coordinates

Many features require plate (or absolute) coordinates rather than pixel coordinates defined with respect to the camera field of view. Prior to recording, all trackers are regularly calibrated to determine the conversion from pixels to absolute coordinates. When recording is complete, stage movements are matched to their video signature in order to convert segmented worms to absolute coordinates (offset by the stage's location).

During recording, every stage movement is logged. When recording has completed, the video is scanned to locate motion frames. Because re-centering the worm causes an abrupt change in both the image background and the worm's location, these changes are simply measured as the pixel variance in the difference between subsequent frames. The Otsu method is used to find an appropriate threshold for delineating stage-movement frames. The number of stage movements and the intervals between them are matched against the log of software-issued stage movement commands. If the match fails (an infrequent event usually caused by worms reaching the boundary of their plate or external factors damaging the recording), the worm and its video are discarded and not used. In our data set, roughly 4% of the videos were discarded due to stage-movement failures.

With the stage movements matched to their video signature, the Otsu threshold is employed once again to compute a local threshold that delineates a more accurate start and end for each individual stage movement. The same algorithm is also employed for the interval at the start of the video until the first stage movement and, similarly, from the last stage movement until the end of the video. With this in place, stage movement frames are discarded and each interval between stage movements is assigned a stage location. Thereafter, each segmented worm is converted to its absolute coordinates on the plate.

Feature Overview

All feature formulas are computed from the worm's segmented contour and skeleton. The skeleton and each side of the contour are scaled down to 49 points for feature computation. Wild-type worms have four quadrants of longitudinal, staggered bodywall muscles²⁷. Each quadrant contains 24 such muscles with the exception of the ventral-left quadrant, which has 23. With a sampling of 49 points, the skeleton and contour sides have a well-defined midpoint. Moreover, since the worm is confined to two dimensions, its bodywall muscles present roughly 24 degrees of freedom (although in practice it seems to be far less⁸). With

49 points we have 2 samples per degree of freedom and, therefore, expect to be sampling above the Nyquist rate for worm posture.

A common notation is used to define the body parts (Supplementary Fig. 4a). The head is controlled by the first four bodywall muscles, per quadrant -- approximately 1/6 the length of the worm²⁸. Similarly, the neck is controlled by the next four bodywall muscles, per quadrant -- also approximately 1/6 the length of the worm. For this reason, we define the head as the first 1/6 of the worm and the neck as the next 1/6 of the worm (skeleton points 1–8 and 9–16, respectively). For symmetry, we define the tail and “hips”, in a similar manner, on the opposite end of the worm. The tail is the last 1/6 of the worm and the hips are defined as the next 1/6 (skeleton points 42–49 and 34–41, respectively). The midbody is defined as the remaining middle 1/3 of the worm (skeleton points 17–33). For some features, the head and tail are further subdivided to extract their tips, the first and last 1/12 of the worm (skeleton points 1–4 and 46–49), respectively.

Frame-by-frame features are represented by top-level histograms and statistics as well as subdivisions exploring their values during forward, backward, and paused states. This is to measure behaviors that depend on the state of motion such as foraging amplitude, which is reduced during reversals in wild-type worms²⁹. Many features are signed to reflect dorsal-ventral orientation, forward-backward trajectory, and other special cases (e.g., eigenworm projection) to capture any asymmetry. Finally, event-style features (coiling, turning, and motion states) are summarized using global and local measures. Global measures include the event frequency, the ratio of time spent within the event to the total experiment time, and a similar measure for the ratio of the distance covered within the event to the total distance traveled by the worm (when available). Local measures include the time spent in every individual event, the distance covered in each event (when available), and both the time and distance covered between each pair of successive events.

Morphology Features

1. Length. Worm length is computed from the segmented skeleton by converting the chain-code pixel length to microns.
2. Widths. Worm width is computed from the segmented skeleton. The head, midbody, and tail widths are measured as the mean of the widths associated with the skeleton points covering their respective sections. These widths are converted to microns.
3. Area. The worm area is computed from the number of pixels within the segmented contour. The sum of the pixels is converted to microns².
4. Area/Length.
5. Midbody Width/Length.

Posture Features

1. Bends. Worm bending is measured using the supplementary angles to the bends formed along the skeleton, with each skeleton point serving as the vertex to its respective bend (Supplementary Fig. 4b). The supplementary angle can also be

expressed as the difference in tangent angles at the skeleton point. The supplementary angle provides an intuitive measurement. Straight, unbent worms have an angle of 0° . Right angles are 90° . And the largest angle theoretically possible, a worm bending back on itself, would measure 180° . The supplementary angle is determined, per skeleton point, using edges $1/12$ the skeleton's chain-code length, in opposing directions, along the skeleton. When insufficient skeleton points are present, the angle remains undefined (i.e., the first and last $1/12$ of the skeleton have no bending angle defined). The mean and standard deviation are measured for each body segment. The angle is signed to provide the bend's dorsal-ventral orientation. When the worm has its ventral side internal to the bend, the bending angle is signed negatively.

2. **Bend Count.** The bend count is a rough measure of the number of bends along the worm. The supplementary skeleton angles are measured during segmentation and signed to reflect their dorsal-ventral orientation. These angles are convolved with a Gaussian filter, $1/12$ the length of the skeleton, with a width defined by the Matlab "gausswin" function's default α of 2.5 and normalized such that the filter integrates to 1, to smooth out any high-frequency changes. The angles are then sequentially checked from head to tail. Every time the angle changes sign or hits 0° , the end of a bend has been found and the count is incremented. Bends found at the start and end of the worm must reflect a segment at least $1/12$ the skeleton length in order to be counted. This ignores small bends at the tip of the head and tail.
3. **Eccentricity.** The eccentricity of the worm's posture is measured using the eccentricity of an equivalent ellipse to the worm's filled contour. The orientation of the major axis for the equivalent ellipse is used in computing the amplitude, wavelength, and track length (described below).
4. **Amplitude.** Worm amplitude is expressed in two forms: a) the maximum amplitude found along the worm body and, b) the ratio of the maximum amplitudes found on opposing sides of the worm body (wherein the smaller of these two amplitudes is used as the numerator). The formula and code originate from the publication "An automated system for measuring parameters of nematode sinusoidal movement"⁶.

The worm skeleton is rotated to the horizontal axis using the orientation of the equivalent ellipse and the skeleton's centroid is positioned at the origin. The maximum amplitude is defined as the maximum y coordinate minus the minimum y coordinate. The amplitude ratio is defined as the maximum positive y coordinate divided by the absolute value of the minimum negative y coordinate. If the amplitude ratio is greater than 1, we use its reciprocal.

5. **Wavelength.** The worm's primary and secondary wavelength are computed by treating the worm's skeleton as a periodic signal. The formula and code originate from the publication "An automated system for measuring parameters of nematode sinusoidal movement"⁶.

The worm's skeleton is rotated as described above for the amplitude. If there are any overlapping skeleton points (the skeleton's x coordinates are not monotonically

increasing or decreasing in sequence -- e.g., the worm is in an S shape) then the shape is rejected, otherwise the Fourier transform computed. The primary wavelength is the wavelength associated with the largest peak in the transformed data. The secondary wavelength is computed as the wavelength associated with the second largest amplitude (as long as it exceeds half the amplitude of the primary wavelength). The wavelength is capped at twice the value of the worm's length. In other words, a worm can never achieve a wavelength more than double its size.

6. **Track Length.** The worm's track length is the range of the skeleton's horizontal projection (as opposed to the skeleton's arc length) after rotating the worm to align it with the horizontal axis. The formula and code originate from the publication "An automated system for measuring parameters of nematode sinusoidal movement"⁶.
7. **Coils.** Worm coiling (touching) events are found by scanning the video frame annotations. During segmentation, every frame that cannot be segmented is annotated with a cause for failure. Two of these annotations reflect coiling events. First, if we find fewer than two sharp ends on the contour (reflecting the head and tail) then the head and/or tail are obscured in a coiling event. Second, if the length between the head and tail on one side of the contour is more than double that of the other side, the worm has either assumed an omega bend or is crossed like a wreath. Empirically, less than 1/5 of a second is a very fast touch and not usually reflective of coiling. Therefore, when a period of unsegmented video frames exceeds 1/5 of a second, and either of the coiling annotations are found, we label the event coiling.
8. **Eigen Projections.** The eigenworm amplitudes are a measure of worm posture. They are the projections onto the first six eigenworms which together account for 97% of the variance in posture. The eigenworms were computed from 15 N2 videos (roughly 3 hours of video, 1/3 of a million frames) as previously described⁸.
Briefly, 48 tangent angles are calculated along the skeleton and rotated to have a mean angle of zero. Principal components analysis is performed on the pooled angle data and we keep the 6 principal components (or eigenworms) that capture the most variance. The first eigenworm roughly corresponds to body curvature. The next two eigenworms are akin to sine and cosine waves encoding the travelling wave during crawling. The fourth eigenworm captures most of the remaining variance at the head and tail. Projected amplitudes are calculated from the posture in each frame. Even for the mutants, the data is always projected onto the N2-derived eigenworms.
9. **Orientation.** The worm's orientation is measured overall (from tail to head) as well as for the head and tail individually. The overall orientation is measured as the angular direction from the tail to the head centroid. The head and tail centroids are computed as the mean of their respective skeleton points.

The head and tail direction are computed by splitting these regions in two, then computing the centroid of each half. The head direction is measured as the angular direction from the its second half (the centroid of points 5–8) to its first half (the

centroid of points 1–4). The tail direction is measured as the angular direction from the its second half (the centroid of points 42–45) to its first half (the centroid of points 46–49).

Motion Features

1. **Velocity.** The worm's velocity is measured at the tip of the head and tail, at the head and tail themselves, and at the midbody. The velocity is composed of two parts, speed and direction (expressed as an angular speed) (Supplementary Fig. 4d). The velocity is signed negatively whenever the respective body part moves towards the tail (as opposed to the head).

The head and tail tips' instantaneous velocity is measured at each frame using a 1/4 second up to a 1/2 second window. For each frame, we search for a start frame 1/4 of a second before and an end frame 1/4 second after to delineate the worm's instantaneous path. If the worm's location is not known within either the start or end frame, we extend the search for a known location up to 1/2 second in either direction. If the worm's location is still missing at either the start or end, the velocity is marked unknown at this point. The speed is defined as the distance between the centroids of the start and end frames (for the respective body parts) divided by the time between both frames. The direction is defined as the angle (between centroids) from the start to the end frame, relative to the worm's overall body angle, divided by the time between both frames. The worm's overall body angle is defined as the mean orientation of the angles, in the tail-to-head direction, between subsequent midbody skeleton points. The body angle is used to sign the velocity. If the head or tail tip's start-to-end angle exceeds 90°, clockwise or anti-clockwise, relative to the overall worm body angle, the motion is towards the tail. In this case both the speed and direction are negatively signed. The head, midbody, and tail velocity are computed identically except they use a 1/2 second up to a 1 second window for choosing their start and end frames.

2. **Motion States.** The worm's forward, backward, and paused motion states attempt to differentiate these event states unambiguously (Supplementary Fig. 4f). Therefore, ambiguous motion has no associated state.

The motion states are computed from the worm's velocity and length (described in the section on "Morphology"). Missing lengths are linearly interpolated between segmented frames. The following filtering criteria were chosen based on human labeling of events within a variety of N2 and mutant videos. The worm is defined in a state of forward motion when a period, more than half a second long, is observed wherein: a) the worm travels at least 5% of its mean length over the entire period; and, b) the worm's speed is at least 5% of its length, per second, in each frame. The worm must maintain this speed almost continuously with permissible interruptions of, at most, a quarter second (this permits quick contradictory movements such as head withdrawal, body contractions, and segmentation noise). The criteria for backward motion is identical except the worm must be moving backwards (the midbody speed must be negatively signed). The worm is defined in a paused state when a period, more than half a second long, is observed wherein the

worm's forward and backward speed do not exceed 2.5% of its length, per second, in each frame. The worm must observe these speed limits almost continuously with permissible interruptions of, at most, a quarter second (once again, this permits quick contradictory movements).

3. **Crawling.** Worm crawling is expressed as both an amplitude and frequency (Supplementary Fig. 4e). We measure these features instantaneously at the head, midbody, and tail. The amplitude and frequency are signed negatively whenever the worm's ventral side is contained within the concave portion of its instantaneous bend.

Crawling is only measured during forward and backward motion states. The worm bend mean angles (described in the section on "Posture") show a roughly periodic signal as the crawling wave travels along the worm's body. This wave can be asymmetric due to differences in dorsal-ventral flexibility or simply because the worm is executing a turn. Moreover the wave dynamics can change abruptly to speed up or slow down. Therefore, the signal is only roughly periodic and we measure its instantaneous properties.

Worm bends are linearly interpolated across unsegmented frames. The motion states criteria (described earlier in this section) guarantee that interpolation is no more than 1/4 of a second long. For each frame, we search both backwards and forwards for a zero crossing in the bend angle mean – the location where the measured body part (head, midbody, or tail) must have hit a flat posture (a supplementary bend angle of 0°). This guarantees that we are observing half a cycle for the waveform. Crawling is bounded between 1/30Hz (a very slow wave that would not resemble crawling) and 1Hz (an impossibly fast wave on agar). If the window between zero crossings is too small, the nearest zero crossing is assumed to be noise and we search for the next available zero crossing in its respective direction. If the window is too big, crawling is marked undefined at the frame. Once an appropriate window has been found, the window is extended in order to center the frame and measure instantaneous crawling by ensuring that the distance on either side to respective zero crossings is identical. If the distances are not identical, the distance of the larger side is used in place of the zero-crossing distance of the smaller side in order to expand the small side and achieve a symmetric window, centered at the frame of interest.

We use a Fourier transform to measure the amplitude and frequency within the window described above. The largest peak within the transform is chosen for the crawling amplitude and frequency. If the troughs on either side of the peak exceed 1/2 its height, the peak is rejected for being unclear and crawling is marked as undefined at the frame. Similarly, if the integral between the troughs is less than half the total integral, the peak is rejected for being weak.

4. **Foraging.** Worm foraging is expressed as both an amplitude and an angular speed (Supplementary Fig. 4g). Foraging is signed negatively whenever it is oriented towards the ventral side. In other words, if the nose is bent ventrally, the amplitude is signed negatively. Similarly, if the nose is moving ventrally, the angular speed is

signed negatively. As a result, the amplitude and angular speed share the same sign roughly only half the time. Foraging is an ambiguous term in previous literature, encompassing both fine movements of the nose as well as larger swings associated with the head. Empirically we have observed that the nose movements are aperiodic while the head swings have periodicity. Therefore, we measure the aperiodic nose movements and term these foraging whereas the head swings are referred to as measures of head crawling (described earlier in this section).

Foraging movements can exceed 6Hz ⁷ and, at 20–30fps, our video frame rates are just high enough to resolve the fastest movements. By contrast, the slowest foraging movements are simply a continuation of the crawling wave and present similar bounds on their dynamics. Therefore, we bound foraging between $1/30\text{Hz}$ (the lower bound used for crawling) and 10Hz .

To measure foraging, we split the head in two (skeleton points 1–4 and 5–8) and measure the angle between these sections. To do so, we measure the mean of the angle between subsequent skeleton points along each section, in the tail-to-head direction. The foraging angle is the difference between the mean of the angles of both sections. In other words, the foraging angle is simply the bend at the head. Missing frames are linearly interpolated, per each skeleton point, for fragments up to 0.2 seconds long (4–6 frames at 20–30fps – twice the upper foraging bound). When larger fragments are missing, foraging is marked undefined. Segmentation of the head at very small time scales can be noisy. Therefore, we smooth the foraging angles by convolving with a Gaussian filter $1/5$ of a second long (for similar reasons to those mentioned in frame interpolation), with a width defined by the Matlab “gausswin” function’s default α of 2.5 and normalized such that the filter integrates to 1.

The foraging amplitude is defined as the largest foraging angle measured, prior to crossing 0° . In other words, the largest nose bend prior to returning to a straight, unbent position. Therefore, the foraging amplitude time series follows a discrete, stair-step pattern. The amplitude is signed negatively whenever the nose points towards the worm’s ventral side. The foraging angular speed is measured as the foraging angle difference between subsequent frames divided by the time between these frames. To center the foraging angular speed at the frame of interest and eliminate noise, each frame is assigned the mean of the angular speed computed between the previous frame and itself and between itself and the next frame. The angular speed is signed negatively whenever its vector points towards the worm’s ventral side.

5. Turns. Omega and epsilon turn events are computed similarly to a previously described method⁹ but using skeleton bends instead of a single head-midbody-tail angle. Omega and epsilon turns are signed negatively whenever the worm’s ventral side is sheltered within the concavity of its midbody bend.

The worm bends (described in the section on “Posture”) are used to find a contiguous sequence of frames (interruptible by coiling and other segmentation failures) wherein a large bend travels from the worm’s head, through its midbody,

to its tail. The worm's body is separated into three equal parts from its head to its tail. The mean supplementary angle is measured along each third. For omega turns, this angle must initially exceed 30° at the first but not the last third of the body (the head but not the tail). The middle third must then exceed 30° . And finally, the last but not the first third of the body must exceed 30° (the tail but not the head). This sequence of a 30° mean supplementary angle, passing continuously along the worm from head to tail, is labeled an omega turn event. Upsilon turns are computed nearly identically but they capture all events that escaped being labeled omega turns, wherein the mean supplementary angle exceeded 15° on one side of the worm (the first or last third of the body) while not exceeding 30° on the opposite end.

Path Features

1. **Range.** The centroid of the worm's entire path is computed. The range is defined as the distance of the worm's midbody from this overall centroid, in each frame (Supplementary Fig. 4h).
2. **Dwelling.** The worm dwelling is computed for the head, midbody, tail, and the entire worm (Supplementary Fig. 4i). The worm's width is assumed to be the mean of its head, midbody, and tail widths across all frames. The skeleton's minimum and maximum location, for the x and y axes, is used to create a rectangular boundary. This boundary is subdivided into a grid wherein each grid square has a diagonal the same length as the worm's width. When skeleton points are present on a grid square, their corresponding body part is computed as dwelling within that square. The dwelling for each grid square is integrated to define the dwelling distribution for each body part. For each body part, untouched grid squares are ignored.
3. **Curvature.** The path curvature is defined as the angle, in radians, of the worm's path divided by the distance it traveled in microns. The curvature is signed to provide the path's dorsal-ventral orientation. When the worm's path curves in the direction of its ventral side, the curvature is signed negatively.

The worm's location is defined as the centroid of its body, with the head and tail removed (points 9–41). We remove the head and tail because their movement can cause large displacements in the worm's centroid. For each frame wherein the worm's location is known, we search for a start frame $1/4$ of a second before and an end frame $1/4$ second after to delineate the worm's instantaneous path. If the worm's location is not known within either the start or end frame, we extend the search for a known location up to $1/2$ second in either direction. If the worm's location is still missing at either the start or end, the path curvature is marked unknown at this point.

With three usable frames, we have an approximation of the start, middle, and end for the worm's instantaneous path curvature. We use the difference in tangent angles between the middle to the end and between the start to the middle. The distance is measured as the integral of the distance traveled, per frame, between the

start and end frames. When a frame is missing, the distance is interpolated using the next available segmented frame. The instantaneous path curvature is then computed as the angle divided by the distance. This path curvature is signed negatively if the angle curves in the direction of the worm's ventral side.

Phenotypic Ontology

The phenotypic ontology attempts to find significant features and reduce our large set of statistical measures to several simple terms. Each ontological term has a prefix indicating whether all significant measurements agree that the feature is greater (+), less (-), or different () than the control. A feature is said to be different than its control whenever the magnitude has no direct meaning (e.g., asymmetry does not translate to a clear description of the measurement being less nor greater than the control) or its measures do not express a simple magnitude (e.g., the strain pauses with greater frequency but spends less time in each paused event). Each term also has a suffix indicating the minimum q-value (significance) found for the term's defining measures (* when $q < 0.05$; ** when $q < 0.01$; *** when $q < 0.001$; and, **** when $q < 0.0001$). The q-value is a p-value replacement that corrects for multiple testing¹⁰. The ontology terms are as follows:

1. Length. The worm's length.
2. Width. The worm's head, midbody, and/or tail width.
3. Area. The worm's area if neither the "Length" nor "Width" were found significant.
4. Proportion. The worm's area/length and/or width/length if neither the "Length", "Width", nor "Area" were found significant.
5. Head Bends. The worm's head bend mean and/or standard deviation.
6. Tail Bends. The worm's tail bend mean and/or standard deviation.
7. Posture Amplitude. The worm's maximum amplitude and/or amplitude ratio.
8. Posture Wavelength. The worm's primary and/or secondary wavelength.
9. Posture Wave. The worm's track length if neither the "Posture Amplitude" nor the "Posture Wavelength" were found significant.
10. Body Bends. The worm's eccentricity, its number of bends, and/or its neck/midbody/hips bend mean and/or standard deviation; only if neither the "Posture Amplitude", "Posture Wavelength", nor "Posture Wave" were found significant.
11. Pose. The worm's eigenworm projections if neither the "Head Bends", "Body Bends", "Tail Bends", "Posture Amplitude", "Posture Wavelength", nor "Posture Wave" were found significant.
12. Coils. The worm's coiling event details.
13. Foraging. The worm's foraging amplitude and/or angular speed.
14. Forward Velocity. The worm's forward (positive) velocity vector.
15. Backward Velocity. The worm's backward (negative) velocity vector.

16. Velocity. The worm's velocity vector magnitude and/or asymmetry if neither the "Forward Velocity" nor "Backward Velocity" were found significant.
17. Head Motion. The worm's head-tip and/or head velocity vectors if neither the "Foraging", "Forward Velocity", nor "Backward Velocity" were found significant.
18. Tail Motion. The worm's tail-tip and/or tail velocity vectors if neither the "Forward Velocity" nor "Backward Velocity" were found significant.
19. Forward Motion. The worm's forward motion event details.
20. Pausing. The worm's pausing event details.
21. Backward Motion. The worm's backward motion event details.
22. Crawling Amplitude. The worm's crawling amplitude.
23. Crawling Frequency. The worm's crawling frequency.
24. Turns. The worm's omega and/or epsilon event details.
25. Path Range. The worm's path range.
26. Path Curvature. The worm's path curvature.
27. Dwelling. The worm's dwelling if its "Pausing" was not found significant.

Feature File Overview

The features are presented within four types of files available online at: <http://wormbehavior.mrc-lmb.cam.ac.uk/>

PDF files provide a visual summary of the data, per strain. CSV files provide a spreadsheet of the data, per strain. And, three types of MAT files are provided to access the strain data and statistics as well as the skeleton, contour, and feature data for each individual experiment, per frame.

The MAT files, per worm, are available for every experiment. To ensure high-quality experimental data, strain collections of experiments and controls were filtered and only include worm videos of at least 20fps, 14–15 minutes long, wherein at least 20% of the frames were segmented. We only include data collected Monday through Saturday, from 8am to 6pm. This resulted in a mean of 24 worms per strain with a minimum of 12 and a standard deviation of 14. Controls were chosen from the filtered N2 data collection by matching the strain collections to controls performed within the same week. This resulted in a mean of 63 controls, per strain collection, with a minimum of 18 and a standard deviation of 29. We examined 100 videos (roughly 2 million frames) from our filtered collection and found that the head was correctly labeled with a mean and standard deviation of $95.17 \pm 17.5\%$ across individual videos and 95.69% of the frames collectively.

Outliers can compress visual details in their corresponding histograms. For this reason, the strain collections underwent one more filtering step prior to inclusion in the PDF files. Experiments were discarded wherein any of the worm data exceeded reasonable bounds of 250 to 2000 microns for length, 25 to 250 microns for width, and/or –1000 to 1000 microns/

seconds for the midbody speed. Outliers were seldom found. Overall, 49 non-control worms were lost from a collection of 7,529 experiments. No strain collection lost more than 2 worms. The N2 collection of controls lost 5 worms from its total of 1,218 experiments. The CSV files and MAT statistical-significance files are available for both the primary quality-filtered data sets and the secondary, outlier-filtered data sets.

Shapiro-Wilk testing (performed using the “swtest” function by Ahmed Ben Saïda) of each feature measure (with corrections for multiple comparisons) showed a maximum q-value of 0.0095 over our collective N2 data set, indicating that, in aggregate, none of the measures are normally distributed. Further testing across all strain collections (which have far lower sampling than the N2 collective) and their controls, indicated a roughly 2:1 ratio of normal to non-normal distributions, rejecting the null hypothesis of normality at a q-value of 0.05. Therefore, we chose to test strain measurements against their controls by using the non-parametric Wilcoxon rank-sum test (with the null hypothesis that both sets of mean values were drawn from the same distribution). In four strains, at least one measure was detected exclusively in either the strain or its control, meaning the measurement was always observed within one set and never in the other (e.g., some strains never perform reversals). When this occurred, we used a Fisher’s exact test to measure the probability that our sets were drawn from the same distribution of observed and unobserved events. Occasionally, features measurements had insufficient values for testing due to low sampling (e.g., omega-turn events), these measures were ignored and their p-value marked as undefined. In total, our 702 measurements were obtained for each of 305 strains in addition to collections of our N2 worms by hour (9am–4pm, with 8am and 5pm discarded due to very low sampling), weekday (Tuesday-Friday, with Monday and Saturday discarded due to very low sampling), and month (January-December). We used False-Discovery Rate (FDR) to correct for nearly 702 measures by 329 groups and transform the p-values to their q-value equivalents¹⁰.

Our unfiltered histograms, presented within individual MAT files, were constructed by choosing standard bin resolutions (widths and centers) that resulted in roughly 10^3 bins, per feature, for our N2 data. When plotting histograms, we use a common formula to downsample the bins. We measure the square root of the total number of data samples contributing to the collective histogram. If this value is less than the number of bins available, the histogram is downsampled to reduce the number of bins to the nearest integer at or below the computed square root. When multiple histograms are plotted together, the smallest common bin size is used to downsample all the histograms to the same bin width and centers.

PDF Files

The PDF (portable document format) files include five sections: a) a table of contents and overview of the results, b) a short summary of the most important features, c) the details for every feature, d) traces of the worm paths, and e) a reference with the experimental methods. Each page uses a color scheme to provide quick visual summaries of its results. All pages display tabs, on the right side, that explain their color scheme. The initial summary page of histograms (page 2) displays an example histogram that acts as a guide to understanding histogram plots and the statistics displayed in their titles. The page formats are as follows:

1. The table of contents details the layout of the PDF file. All feature measures are shown alongside their minimum q-value and a page number for details. The table of contents page also shows an overview with the experiment annotation and its phenotypic ontology (see the section titled “Phenotypic Ontology”).
2. There are three summary pages. These pages show important feature histograms, with the collective experiments in color and their controls in gray. The background color, for the histogram plots, indicates the minimum q-value significance for the plotted feature. The title of each plot provides several statistical measures for the experiment and control collections. An example histogram, at the beginning of the first summary page, provides a reference to interpret the aforementioned statistical measures. Significant measures, with $q < 0.05$, are marked in bold font within the plot title.

The crawling frequency, worm velocity, foraging speed, all event features, path range, and dwelling are shown on a pseudo log-value scale to improve readability within their small summary histograms. This pseudo log-value scale is achieved by taking the magnitude of the data values (to avoid complex numbers resulting from the logarithms of any negative numbers), translating the magnitude by 1 (to avoid the logarithms of any values less than 1, which would invert the sign of the data), taking the logarithm, then re-signing the formerly negative data values.

3. The detail pages present a detailed view of the histograms for every feature. They follow a similar format to the summary pages except that they never use a log scale for feature values. The title of each plot provides a large set of statistical measures. The control values are shown between square brackets. The statistical values include: a) the number of worms providing measurements (“WORMS”); b) the number of measurements sampled for the collection of worms (“SAMPLES”); c) the mean of the data (“ALL”) alongside the SEM and, when the data is signed, the means for the absolute data values (ABS), positive data values only (“POS”), and negative data values only (“NEG”) alongside their SEMs as well; d) the p-value results using Wilcoxon rank-sum testing and q-value results using False Discovery Rate correction (for multiple tests across 329 strain collections by 702 feature measurements), both labeled accordingly (respectively “p” and “q”); e) event features also display their mean frequency (“FREQ”), the mean percentage of time spent in the event relative to the total experiment time (“TIME”), and, when available, the mean percentage of distance traveled during the event relative to the total distance covered during the experiment (“DIST”).

Features that have motion-state subdivisions are shown with an additional view wherein all motion-state histograms, and their integral histogram, are shown on the same plot. This allows one to quickly distinguish behaviors dependent on the motion state. Event features have an additional view wherein event and inter-event measures are plotted on a log-probability scale to make outlying events more visible.

4. The path trace pages display the paths for the worms’ head, midbody, and tail and heatmaps for the midbody speed and foraging amplitude. Pages with the head,

midbody, and tail include a tab, on the right side, to interpret the color associated with each body part. Pages with heatmaps include a tab, on the right side, to interpret the color gradient. On the path trace plots, the start and end of each path is denoted by a gray and black worm, respectively. Moreover, on each plot, the locations for coiling events are marked by a “+” and those for omega turns are marked by an “x”. Body part plots use transparency to roughly indicate dwelling through color opacity.

The first page of each path trace shows a collection of up to 24 worms (when available) overlaid for both the experiment and control collections, at the same scale. These overlays provide a quick view of features such as relative path sizes, food leaving behaviors, and the relative locations for coiling events and omega turns. When more than 24 worms are available we sort the worms by date, then choose 24 from the first to the last experiment at regular intervals. The paths are rotated to align their longest axis vertically, and then centered using the minimum and maximum x and y path values, per worm.

The next page of the path traces shows each collection of 24 paths on the same plot, ordered roughly from largest to smallest, spaced out to avoid any overlay. The experiments and their controls use independent scales. This ordered plot provides a quick view to distinguish salient characteristics of experiment versus control paths (e.g., bordering at the edge of the food lawn).

The subsequent pages for each path trace show the 24 individual worm paths, for the experiments and their controls, without rotation, sorted by date.

5. The method pages provide a reference for the details of our methodology.

CSV Files

The CSV (comma separated value) files are compatible with popular spreadsheet programs (e.g., Microsoft Excel, Apple iWork Numbers, OpenOffice, etc.). Each experimental collection is accompanied by four CSV files presenting the data and statistics for all morphology (<filename>.morphology.csv), posture (<filename>.posture.csv), motion (<filename>.motion.csv), and path features (<filename>.path.csv). The CSV files present the strain, genotype, and date for the experimental strain and control worms. The mean and standard deviation are presented for each feature measure, per worm and for the collection of experiments and controls. The p and q-values are presented for the strain as a whole (the null hypothesis is that experiment and control worms are drawn from the same distribution) and for each feature measure individually. These p and q values are shown for both the non-parametric Wilcoxon rank-sum test and the normal-distribution Student's *t*-test (unpaired samples with unequal variance). The Shapiro-Wilk test for normality (with associated p and q values) is also shown for each measure. Correction for multiple testing (the q-values) was performed over our entire set of 329 groups of strain collections by 702 measures. For the Shapiro-Wilk normality test, correction for multiple comparisons included an additional 329 group-specific controls by 702 measures.

MAT Files

Each experiment is represented in a MAT, HDF5-formatted file (Hierarchical Data Format Version 5 – an open, portable, file format with significant software support). HDF5 files are supported by most popular programming languages including Matlab, Octave (a free alternative to Matlab), R, Java, C/C++, Python, and many other environments. These experiment files contain the time-series feature data for an individual worm. Additionally, each strain collection of experiments and their collection of controls are also represented in a single HDF5, MAT file. These strain files contain histogram representations and summary statistics (but not significance) for the collective experiments. Finally, the statistical significance, for our entire collection of mutants, is presented in a single HDF5, MAT file.

The first two MAT file types, individual experiments and strain collections, share a similar format. The individual experiment files present the feature data as a time series. They also include the full skeleton and the centroid of the contour, per frame, permitting novel feature computations. The strain collections present the data in summary and in histograms. The format for both file types is two top-level structs, “info” (“wormInfo” for the strain collections) and “worm”, which contain the experimental annotation and data, respectively.

The “info” struct contains the experimental annotation. For the strain collections, the “info” from each experiment is collected into an array of structs called “wormInfo”. Both variables share the same format with the following subfields:

1. wt2. The Worm Tracker 2.0 version information.
2. video. The video information. The video “length” is presented as both “frames” and “time”. The video “resolution” is in “fps” (frames/seconds), pixel “height” and “width”, the ratio of “micronsPerPixel”, and the codec’s “fourcc” identifier. The video frame “annotations” are presented for all “frames” with a “reference” specifying the annotation’s numerical “id”, the “function” it originated from, and a “message” describing the meaning of the annotation.
3. experiment. The experiment information. The “worm” information is presented for its “genotype”, “gene”, “allele”, “strain”, “chromosome”, “sex”, “age”, the “habituation” time prior to recording, the location of its “ventralSide” in the video (clockwise or anti-clockwise from the head), the “agarSide” of its body (the body side touching the agar), and any other worm “annotations”. The “environment” information is presented for the experiment conditions including the “timestamp” when the experiment was performed, the “arena” used to contain the worm (always a low-peptone NGM plate for the data presented here), the “food” used (e.g., OP50 *E. coli*), the “temperature”, the peak wavelength of the “illumination”, any “chemicals” used, the “tracker” on which the experiment was performed (a numerical ID from 1 to 8), and any other environmental “annotations”.
4. files. The name and location for the analyzed files. Each experiment is represented in a “video” file, “vignette” file (a correction for video vignetting), “info” file (with tracking information, e.g., the microns/pixels), a file with the log of “stage” movements, and the “computer” and “directory” where these files can be found.

5. lab. The lab information where the experiment was performed. The lab is represented by its “name”, the “address” of the lab, the “experimenter” who performed the experiment, and any other lab-related “annotations”.

The “worm” struct contains experimental data. The individual experiments contain the full time series of data along with the worm’s skeleton and the centroid of its contour, per frame. The strain collections contain summary data and histograms in place of the time-series data. Both files share a similar initial format with the following subfields:

1. morphology. The morphology features. The morphology is represented by the worm’s “length”, its “width” at various body locations, the “area” within its contour, the “widthPerLength”, and the “areaPerLength”.
2. posture. The posture features. The worm’s posture is represented by its bend count in “kinks”, measures of the “bends” at various body locations (computed as both a “mean” and standard deviation, “stdDev”), its “max” “amplitude” and its “ratio” on either side, its “primary” and “secondary” “wavelength”, its “trackLength”, its “eccentricity”, its “coils”, the orientation “directions” of various body parts, and its six “eigenProjections”. Individual experiment files also contain the “skeleton” “x” and “y” coordinates, per frame.
3. locomotion. The motion features. Worm motion states are represented by “forward”, “backward”, and “paused” events, the “speed” and angular “direction” of the “velocity” for various body parts, the “amplitude” and “frequency” of the crawling “bends” for various body parts, as well as the “foraging” “bends” which are measured in an “amplitude” and “angleSpeed”, and the “turns” associated with “omega” and “upsilon” events. Individual experiment files also contain a “motion” state “mode” with values distinguishing forward (1), backward (−1), and paused (0) states, per frame.
4. path. The path features. The path is represented by its “range”, “curvature”, and the dwelling “duration” for various body parts. Individual experiment files also contain the “x” and “y” “coordinates” of the contour’s centroid. Moreover, the individual experiment files present the “duration” as an “arena” with a “height”, “width”, and the “min” and “max” values for the “x” and “y” axes of the arena. The arena can be transformed to a matrix using the given height and width. The duration of the worm and body parts are represented as an array of “times” spent at the “indices” of the arena matrix.

All events are represented by their “frequency” and either their “timeRatio” (the ratio of time in the event type to the total experiment time) or, if the worm can travel during the event, the “ratio.time” (equivalent to “timeRatio”) and “ratio.distance” (the ratio of the distance covered in the event type to the total distance traveled during the experiment). The individual experiment files represent each event as “frames” with a “start” frame, “end” frame, the “time” spent in this event instance, the “distance” traveled during this event instance (when available), the “interTime” till the next event, and the “interDistance” traveled till the next event. The strain collection files summarize these fields, excluding the individual “frames” and their “start” and “end”.

The strain collection files present the data for each feature within a “histogram” (as opposed to the individual experiment files which simply use a time-series array of values). Furthermore, when a feature can be subdivided by motion state, sub histograms are included for the “forward”, “backward”, and “paused” states. All histograms contain the “PDF” (probability distribution function) for each of their “bins” (centered at the associated feature’s values). All histograms also contain the “resolution” (width) of their bins, whether or not there “isZeroBin” (would one of the bins be centered at 0?), and whether or not the feature “isSigned” (can the feature values be negative?).

Finally, the strain collection files present their data in three types of fields: a) individually as the “data” per experiment, b) summarized over the “sets” of experiments and, c) aggregated in “allData” as if we ran one giant experiment instead of our sets. In other words, “sets” weights each experiment identically whereas “allData” weights every frame, across all experiments, identically. The data is always represented as both a “mean” and “stdDev” (standard deviation). The mean and standard deviation are always computed for “all” the data. When the data is signed, the mean and standard deviation are also computed for the data’s “abs” (absolute value), “pos” (only the positive values), and “neg” (only the negative values). The format for the three types of data is as follows:

1. data. The individual data for every experiment is presented in arrays (in the same order as the “wormInfo” experiment annotations). The array data presents each experiment’s individual “mean”, “stdDev”, the number of “samples” measured, and the experiment’s data “counts” for each one of the histogram’s “bins”.
2. sets. The data for the set of experiments is presented as the “mean”, “stdDev”, and “samples” (the number of experiments) of the collected set.
3. allData. The aggregate of all data measurements, as if the collection of videos were instead one long, giant video, is presented as a “mean”, “stdDev”, the total “samples” (the total number of frames wherein the data was measured), and the aggregate of “counts” for each one of the histogram’s bins.

Statistical Significance MAT File

The statistical significance for all strains is collected into a single MAT file. This file contains three top-level structs with information for both the “worm” and “control” collections as well as the “dataInfo” necessary to interpret the included matrices of data. The matrices are organized as rows of strains and columns of feature measures. The “worm” struct has the following subfields:

1. info. The worm information for each strain collection presented as their “strain”, “genotype”, “gene”, and “allele”.
2. stats. The statistics for each strain collection presented, for every feature measure, as their “mean”, “stdDev” (standard deviation), “samples” (the number of worms providing a measurement for the feature – e.g., not all worms execute omega turns), and “zScore” relative to the control (a simple normalization to the control -- note that the collection of N2 controls has no zScore). Measurements exclusively found in the experimental group have a zScore of infinity and those found exclusively

found in the control are -infinity. Furthermore, we include Shapiro-Wilk tests of data normality, per measure, in “pNormal” and correction for multiple testing, using their False-Discovery rate q-value replacements, in “qNormal”. The q-values are computed across all measures per “strain” and their associated controls (roughly 1404 tests) and across “all” strain and control measures collectively (roughly 329 by 1404 tests).

3. sig. The statistical significance for each strain collection is presented, for every feature measure, as their “pTValue” (Student’s *t*-test p-value, unpaired samples with unequal variance) and “pWValue” (Wilcoxon rank-sum test p-value). The “qTValue” and “qWValue” represent the False-Discovery rate q-value replacements for the “pTValue” and “pWValue” respectively. The q-values are computed across all measures per “strain” (roughly 702 tests) and across “all” strains and measures collectively (roughly 329 by 702 tests). The collection of N2s has no associated significance.

The “control” struct contains the control “stats” in an identical format to the “worm” struct “stats”, but without the “zScores”.

The “dataInfo” provides information for each column of the feature measure matrices used in the “worm” and “control” structs. Each feature measure has a “name”, a “unit” of measurement, titles for three possible subdivisions (“title1”, “title2”, and “title3” – the title of the feature itself, its motion state, and its signed subdivision), helpful indexed offsets for these titles (“title1I”, “title2I”, and “title3I”), an associated struct “field” to locate the feature in our other MAT files, the corresponding “index” for the struct field (e.g., the six eigenworm projections are represented in a field, as a 6-element array), “isMain” (is this the main feature as opposed to a subdivision of a main feature?), the feature “category” (morphology “m”, posture “s”, motion “l”, path “p”), the feature “type” (simple data “s”, motion data “m”, event summary data “d”, event data “e”, inter-event data “i”), the feature “subtype” (none “n”, forward motion state “f”, backward motion state “b”, paused state “p”, event-time data “t”, event-distance data “d”, event-frequency data “h”), and information regarding the feature’s “sign” (the feature is signed “s”, unsigned “u”, is the absolute value of the data “a”, contains only positive data values “p”, contains only negative data values “n”).

Supplementary Material

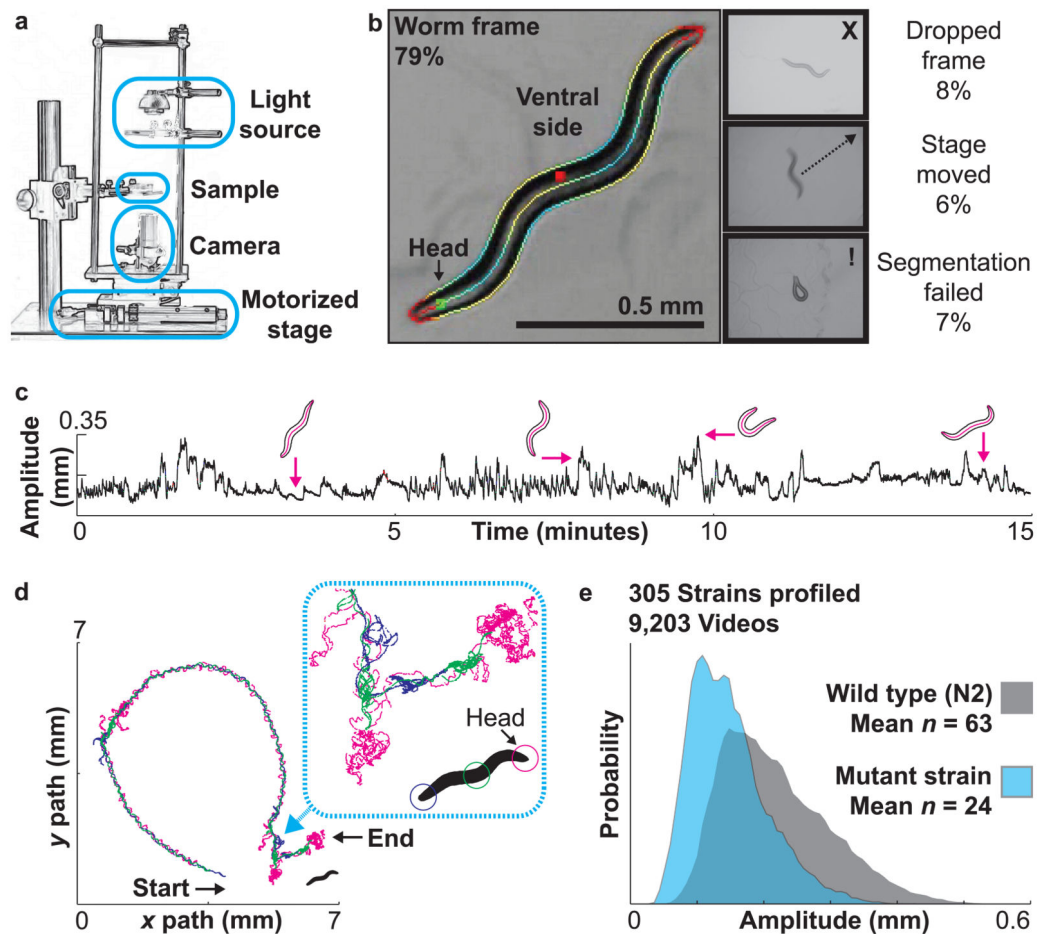
Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Cronin and P. Sternberg for providing the code from their single-worm tracker which serves as the basis for elements of our WT2 analysis GUI and phenotypic features cited from their publication. J. Lasenby and N. Kingsbury for guidance in computer vision. A. Deonarine for help in producing the clustering figure. R. Samworth, S. Chavali, G. Chalancon, S. Teichmann, and M. Babu and for guidance in statistics and bioinformatics techniques. This project was funded by the MRC (MC-A022- 5PB91) the NIH/NIDA (DA018341), a Gates Cambridge Scholarship (to E.Y.) and a long-term fellowship from the Human Frontier Science Program (to A.E.X.B).

References

1. Ramani AK, et al. *Cell*. 2012; 148:792–802. [PubMed: 22341449]
2. Houle D, Govindaraju DR, Omholt S. *Nature reviews genetics*. 2010; 11:855–866.
3. Husson SJ, Costa WS, Schmitt C, Gottschalk A. Keeping track of worm trackers. *WormBook: the online review of C. elegans biology*. 2012:1–17.10.1895/wormbook.1.156.1
4. Hart, AC. The C. elegans Research Community. *WormBook: the online review of C. elegans biology*. 2006. p. 1-67.
5. Geng W, Cosman P, Berry CC, Feng Z, Schafer WR. *IEEE transactions on bio-medical engineering*. 2004; 51:1811–1820. [PubMed: 15490828]
6. Cronin CJ, et al. *BMC Genet*. 2005; 6:5. [PubMed: 15698479]
7. Huang KM, Cosman P, Schafer WR. *Journal of neuroscience methods*. 2008; 171:153–164. [PubMed: 18342950]
8. Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS. *PLoS computational biology*. 2008; 4:e1000028. [PubMed: 18389066]
9. Huang KM, Cosman P, Schafer WR. *Journal of neuroscience methods*. 2006; 158:323–336. [PubMed: 16839609]
10. Storey JD. *Journal of the royal statistical society: series B*. 2002; 64:479–498.
11. Starich TA, Xu J, Skerrett IM, Nicholson BJ, Shaw JE. *Neural development*. 2009; 4:16. [PubMed: 19432959]
12. Yeh E, et al. *PLoS biology*. 2008; 6:e55. [PubMed: 18336069]
13. Boulton T, et al. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:18590–18595. [PubMed: 19020092]
14. Kindt KS, et al. *Nature neuroscience*. 2007; 10:568–577. [PubMed: 17450139]
15. Jose AM, Bany IA, Chase DL, Koelle MR. *Genetics*. 2007; 175:93–105. [PubMed: 17057248]
16. Feng Z, et al. *Cell*. 2006; 127:621–633. [PubMed: 17081982]
17. Ghosh R, Mohammadi A, Kruglyak L, Ryu WS. *BMC biology*. 2012; 10:85. [PubMed: 23114012]
18. Brown AE, Yemini EI, Grundy LJ, Jucikas T, Schafer WR. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:791–796. [PubMed: 23267063]
19. Yemini E, Kerr RA, Schafer WR. *Cold Spring Harbor protocols*. 2011; 2011:1475–1479. [PubMed: 22135667]
20. Tsai CA, Chen JJ. *Bioinformatics*. 2009; 25:897–903. [PubMed: 19254923]
21. Zhao B, Khare P, Feldman L, Dent JA. *The Journal of neuroscience*. 2003; 23:5319–5328. [PubMed: 12832557]
22. McGrath PT, et al. *Neuron*. 2009; 61:692–699. [PubMed: 19285466]
23. Weber KP, et al. *PloS one*. 2010; 5:e13922. [PubMed: 21085631]
24. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974; 77:71–94. [PubMed: 4366476]
25. Otsu N. *Automatica*. 1975; 11:23–27.
26. Freeman H. *Electronic Computers, IRE Transactions on EC-10*. 1961:260–268.
27. Sulston JE, Horvitz HR. *Dev Biol*. 1977; 56:110–156. [PubMed: 838129]
28. White J, Southgate E, Thomson J, Brenner S. *Phil Trans R Soc Lond (Biol)*. 1986; 314:1–340. [PubMed: 22462104]
29. Alkema MJ, Hunter-Ensor M, Ringstad N, Horvitz HR. *Neuron*. 2005; 46:247–260. [PubMed: 15848803]

**Figure 1.**

The Worm Tracker 2.0 (WT2) system and phenomic database (<http://wormbehavior.mrc-lmb.cam.ac.uk/>). (a) Schematic of the tracking hardware. (b) Videos are segmented to extract the worm contour and skeleton for phenotypic analysis. Frames from which the worm could not be extracted are annotated with a color-coded border. The border indicates whether the frame was dropped (due to the processor's inability to keep up with incoming video frames), a stage movement occurred (blurring the worm image), or worm extraction failed (due to an unrecognized shape); features were not computed from such frames. Numbers indicate percentage in each category (mean \pm standard deviation). (c) An example of a single-worm behavioral time series measuring maximum amplitude, defined relative to the major axis of the best-fit ellipse (see Methods for details). Several shapes are shown with their corresponding amplitude. (d) An example of a single-worm path (the same worm as in panel c). (e) An example of a feature histogram measuring maximum amplitude (for the strain presented in panels c and d).

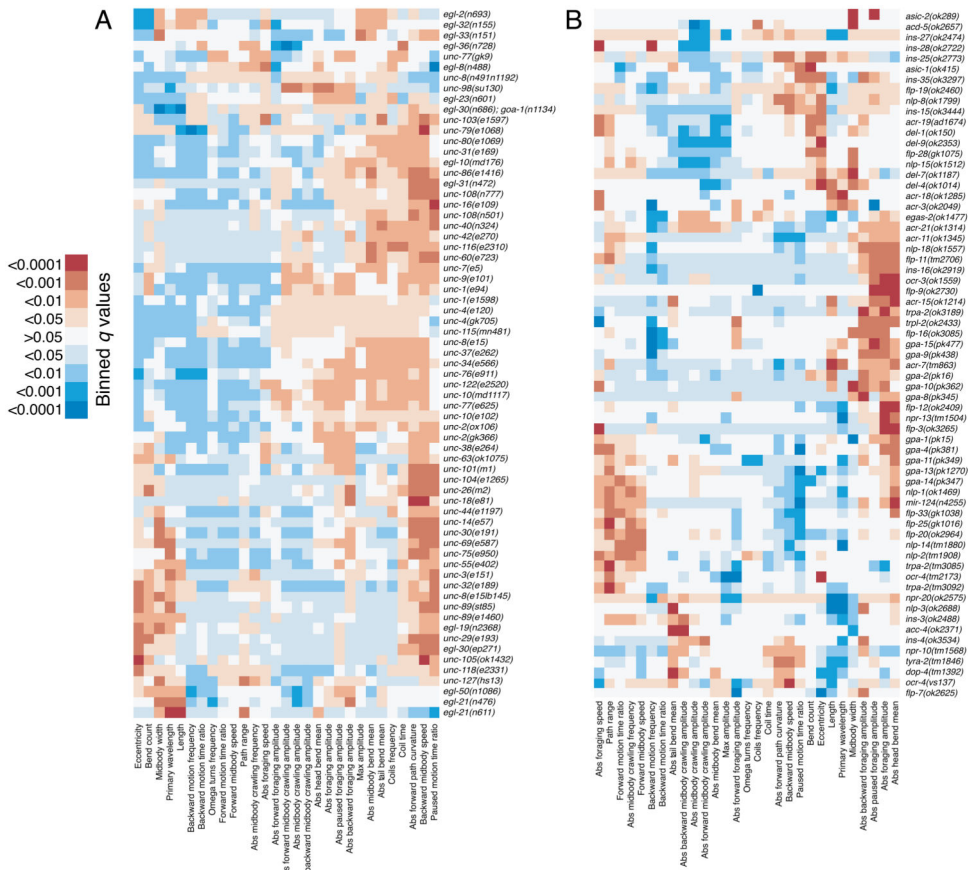


Figure 2. Phenotypic summaries using selected features for subsets of strains. Colors in the heat maps are used to indicate the q-value for each feature for the comparison between each of the mutant strains listed on the right and the N2 reference data. Red values indicate features that have a significantly higher value in the mutant while blue indicates significantly lower values. Genes and features were both hierarchically clustered for easier comparison. (a) A subset of mutant strains with previously known locomotion phenotypes. (b) A subset of mutant strains with no previously reported locomotion phenotype.