A database of *Caenorhabditis elegans* behavioral phenotypes

Eviatar Yemini, Tadas Jucikas, Laura J Grundy, André E X Brown & William R Schafer

Using low-cost automated tracking microscopes, we have generated a behavioral database for 305 *Caenorhabditis elegans* strains, including 76 mutants with no previously described phenotype. The growing database currently consists of 9,203 short videos segmented to extract behavior and morphology features, and these videos and feature data 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.

A 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 has given insight into diverse neural functions. 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 RNAi, though most knockdowns do detectably reduce fitness over several generations[1]. 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.

For this phenotyping gap to be closed, it is desirable to have a system that is capable of measuring phenotypes both extensively and intensively[2]. 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 with single-worm trackers that follow freely behaving worms and that use a motorized stage to keep the worm in the camera's field of view[3].

Here we report extensive and intensive behavioral phenotypes for 9,203 individuals representing 305 strains of *C. elegans* (**Supplementary Table 1**). Included in this data set are most canonical *unc* and *egl* mutants as well as representative knockouts from receptor, channel and neuropeptide gene families that are likely to have roles in nervous system function. The database contains 76 mutants with no previously characterized phenotype, 14 genes with multiple alleles and 10 double- or triple-mutant combinations (the majority involving mutants that are also represented individually). The videos are accessible online at the video-sharing website YouTube on the channel "C.elegans Behavioural Database." The data are also available online (http://wormbehavior.mrc-lmb.cam.ac.uk/) with various degrees of processing, from the skeleton and outline coordinates to the time series of extracted features, the histograms of these features and an in-depth view of their summary statistics. For computational researchers, the database is a rich source of processed

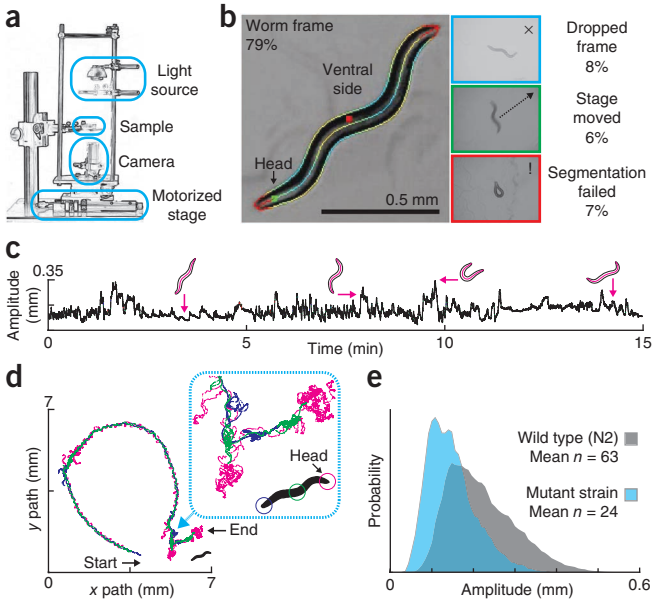

Figure 1 | The Worm Tracker 2.0 (WT2) system and phenomic database (http://wormbehavior.mrc-lmb.cam.ac.uk/). (**a**) Schematic of the tracking hardware. (**b**) Videos are segmented to extract the worm contour and skeleton for phenotypic analysis. Frames from which contour and skeleton cannot be extracted are annotated with a color-coded border. The border indicates whether the frame was dropped (because of the processor's inability to keep up with incoming video frames), a stage movement occurred (blurring the worm image) or worm extraction failed (owing to an unrecognized shape); features are not computed from such frames. Numbers indicate the percentage of frames in each category. (**c**) Single-worm behavioral time series measuring maximum amplitude, defined relative to the major axis of the best-fit ellipse (Online Methods). Several shapes are shown with their corresponding amplitudes. (**d**) Single-worm path (same worm as in **c**). (**e**) Feature histogram measuring maximum amplitude (for the strain presented in **c** and **d**).

Medical Research Council (MRC) Laboratory of Molecular Biology, Cambridge, UK. Correspondence should be addressed to A.E.X.B. (abrown@mrc-lmb.cam.ac.uk) or W.R.S. (wschafer@mrc-lmb.cam.ac.uk).

RECEIVED 7 JANUARY; ACCEPTED 1 JUNE; PUBLISHED ONLINE 14 JULY 2013; DOI:10.1038/NMETH.2560

measures and raw data that can be used to develop new algorithms for segmentation, behavioral quantification and bioinformatic approaches that 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 with Worm Tracker 2.0 (WT2), which uses 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, and the dorsal-ventral orientation was obtained from user input. We further reduced contour and skeleton data by extracting relevant features[4] using automated methods developed for use in previously described worm trackers[5–9]. We developed several new algorithms for feature extraction to quantify parameters related to motion state, crawling, foraging, dwelling and exploratory range (**Supplementary Note**). 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 video 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, such as the frequency, time spent in execution and distance covered during the event. All together, the permutation of these measures yielded 702 distinct feature measurements. By running eight tracking units in tandem, we could achieve a throughput of approximately 128 recordings (8 recordings each of 15 strains plus 18 wild-type controls) per day.

The data from each individual are 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 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 ways to verify the fidelity of the features. Furthermore, aggregate measurements and statistics per strain are available directly on the

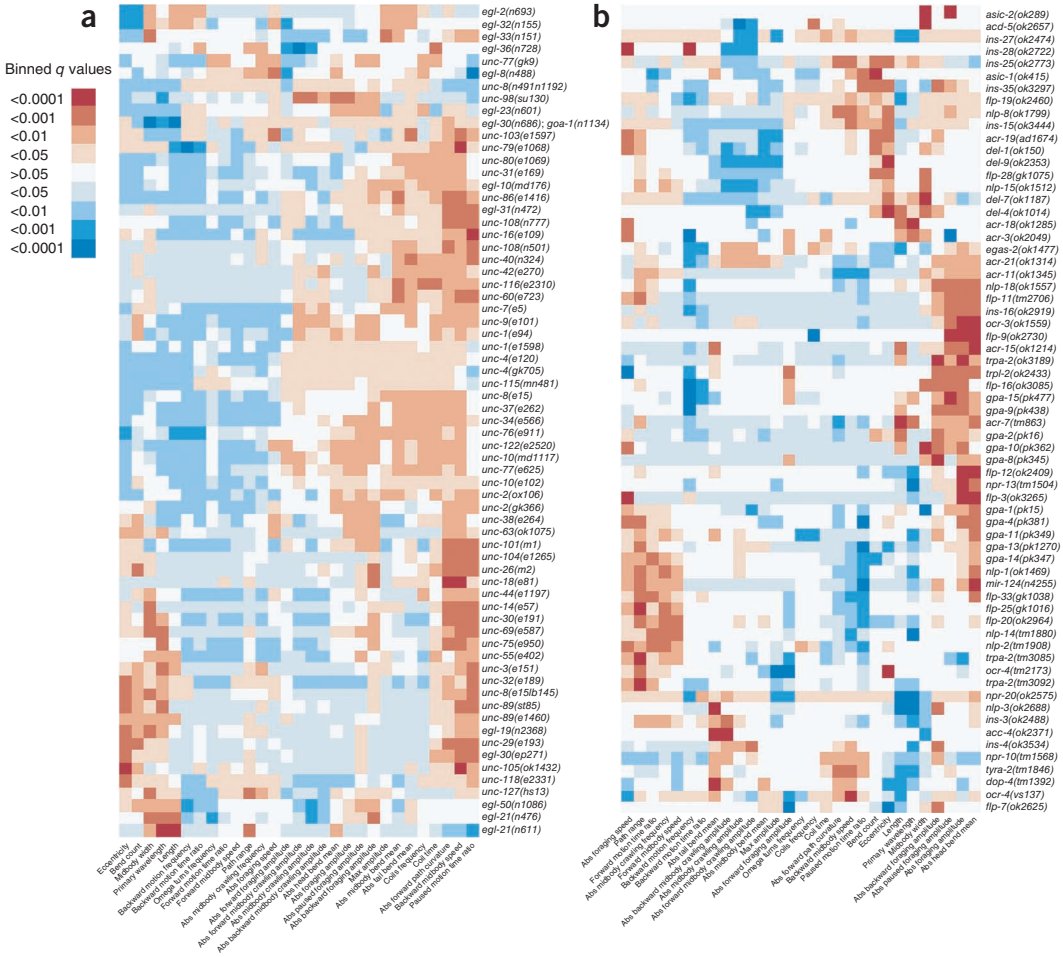

Figure 2 | Phenotypic summaries using selected features for subsets of strains. Colors in the heat maps are used to indicate the *q* value for each feature for the comparison between each of the mutant strains listed on the right and the N2 reference data. Red values indicate features that have a significantly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

website, presented graphically in PDF files, in HDF5-compatible MAT files, and as spreadsheet-compatible CSV files (Online Methods). Inclusion of the skeleton data 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 Fig. 3**), each mutant strain was compared to an N2 control data set of videos collected within a 2-week window centered around the date the mutant data were collected (meaning within 7 d before or after the mutant recording). 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 (ref. 10). All strains were significantly different from wild type at $q \leq 0.05$, 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 (**Fig. 2a**), the measured locomotion features serve as a sensitive fingerprint for phenotypic similarity. Loss-of-function alleles of the same gene (for example, of *unc-2*, *egl-21*, *unc-4*, *unc-108* and *unc-89*) exhibit highly similar phenotypic signatures, as do mutants affecting genes encoding components of known molecular complexes (for example, *unc-7*/*unc-9* (ref. 11), *unc-79*/*unc-80* (ref. 12) and *unc-38*/*unc-63* (ref. 13)). It is reasonable to suppose that the signatures of genetically uncharacterized mutants (such as *egl-31* and *egl-33*) might provide insight into the molecular or cellular functions of the mutant gene. Detailed clustering results are in **Supplementary Figure 5**.

With regard to phenotypes identified in previously uncharacterized knockout strains (**Fig. 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]. 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 Fig. 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 Fig. 6**), including the TRPV channel OCR-4 (ref. 15) and the TRPC channel TRP-2 (ref. 16). 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 skeleton data will make it possible for investigators to develop their own approaches to studying the behavior of particular mutant strains.

We intend to add more strains to the phenotype database and to increase the richness of the data available for each genotype. One approach would be to record each strain under a variety of 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 approach[17]. It should also 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 motifs[18]. 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

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

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 useful discussions; and R. Samworth, S. Chavali, G. Chalancon, S. Teichmann and M. Babu for guidance in statistics and bioinformatics techniques. This project was funded by grants to W.R.S. from the MRC (MC-A022-5PB91) and the National Institute on Drug Abuse, US National Institutes of Health (DA018341), a Gates Cambridge Scholarship (to E.Y.) and a long-term fellowship from the Human Frontier Science Program (to A.E.X.B.).

AUTHOR CONTRIBUTIONS

E.Y. created the WT2 system (hardware and software), analyzed data and wrote the paper. T.J. developed the WT2 analysis pipeline and built the online database. 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.

1. Ramani, A.K. *et al. Cell* **148**, 792–802 (2012).
2. Houle, D., Govindaraju, D.R. & Omholt, S. *Nat. Rev. Genet.* **11**, 855–866 (2010).
3. Husson, S.J., Costa, W.S., Schmitt, C. & Gottschalk, A. in *WormBook* (ed. Hobert, O.), doi:10.1895/wormbook.1.156.1 (2012).
4. Hart, A.C. in *WormBook* (ed. Ambros, V.), doi:10.1895/wormbook.1.87.1 (2006).
5. Geng, W., Cosman, P., Berry, C.C., Feng, Z. & Schafer, W.R. *IEEE Trans. Biomed. Eng.* **51**, 1811–1820 (2004).
6. Cronin, C.J. *et al. BMC Genet.* **6**, 5 (2005).
7. Huang, K.M., Cosman, P. & Schafer, W.R. *J. Neurosci. Methods* **171**, 153–164 (2008).
8. Stephens, G.J., Johnson-Kerner, B., Bialek, W. & Ryu, W.S. *PLOS Comput. Biol.* **4**, e1000028 (2008).
9. Huang, K.M., Cosman, P. & Schafer, W.R. *J. Neurosci. Methods* **158**, 323–336 (2006).
10. Storey, J.D. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).
11. Starich, T.A., Xu, J., Skerrett, I.M., Nicholson, B.J. & Shaw, J.E. *Neural Dev.* **4**, 16 (2009).
12. Yeh, E. *et al. PLoS Biol.* **6**, e55 (2008).
13. Boulin, T. *et al. Proc. Natl. Acad. Sci. USA* **105**, 18590–18595 (2008).
14. Kindt, K.S. *et al. Nat. Neurosci.* **10**, 568–577 (2007).
15. Jose, A.M., Bany, I.A., Chase, D.L. & Koelle, M.R. *Genetics* **175**, 93–105 (2007).
16. Feng, Z. *et al. Cell* **127**, 621–633 (2006).
17. Ghosh, R., Mohammadi, A., Kruglyak, L. & Ryu, W.S. *BMC Biol.* **10**, 85 (2012).
18. Brown, A.E., Yemini, E.I., Grundy, L.J., Jucikas, T. & Schafer, W.R. *Proc. Natl. Acad. Sci. USA* **110**, 791–796 (2013).

ONLINE METHODS

Collection of worm video data. Worms were maintained under strictly controlled conditions up until the point of tracking[19]. To ensure sufficient phenotypic sampling, we filmed at least 20 young-adult hermaphrodites per strain for 15 min that were spontaneously behaving on food. The camera magnification was set to between 3.5 and 4.5 µm per pixel (a corresponding field of view (FOV) of approximately 2.5 × 2 mm^2 at 640 × 480 resolution) and the frame rate was set to 20–30 frames per second to ensure a high-resolution analysis. We observed a 30-min wait, before 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 1**.

To avoid potential room conditions that may bias measurement, we randomized recording, 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), we ignored videos with less than 3 min worth of segmented frames 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 were found to never 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), which represented a broad range of mutants (including several nearly motionless UNCs). Many of these include early videos that suffered multiple dropped frames and poor imaging conditions that were later improved. We found that the head was correctly labeled with a mean and s.d. of 94.39% ± 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 we used Fisher's exact test for measurements found exclusively in the experiment or control groups. Because all features were measured using automated algorithms, scoring was effectively blind. We controlled false discovery rate across these comparisons, converting P values to their q equivalents[10]. 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 $q \leq 0.05$. To ensure this was not an artifact of our methodology, we chose a second method to verify our result. As sparse sampling led to a difficulty in assessing normality, and as 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 values[20]. In four strains, at least one measure was detected exclusively in either the strain or its control, meaning that the measurement was always observed in 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. For all other groups, we ran 10,001 permutations. We corrected the P values, controlling the false discovery rate, to their q equivalents. Once again, every group was found to be significantly different from the wild type at $q \leq 0.05$, bearing out the results of our primary methodology (**Supplementary Table 3**).

Reproducibility and sensitivity of measurements. We performed several 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 h 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 $\tau = 19$ min ($R^2 = 0.96$). Because 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 $\tau = 19$ min ($R^2 = 0.71$; **Supplementary Fig. 2b**), whereas amplitude ("Absolute Midbody Amplitude") decayed with a longer time constant of $\tau = 48$ min ($R^2 = 0.59$; **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 3 years, from 2009 to 2012. Animals were tracked from January to December, from Tuesday through Friday, from 9 a.m. to 5 p.m. We chose six representative measures to analyze variability: length, forward speed ("Positive Midbody Speed"), foraging amplitude ("Absolute Foraging Amplitude"), reversals ("Backward Motion Frequency"), coiling frequency and exploratory range (**Supplementary Figs. 2d** and **3**). To assess the influence of these factors (hour, day and month), we performed one-way ANOVAs with Bonferroni correction for 18 tests and checked for significance at $P \leq 0.05$. The tested groups showed a mixture of normality and non-normality when we used a Shapiro-Wilk test with correction for multiple comparisons. Therefore, to avoid the assumption of normality, we also performed Kruskal-Wallis tests with the same correction and α, which resulted in identical significance. Because 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) among animals tracked throughout the day. We measured changes in length and midbody width, over 2 h, among the previously mentioned group of 25 young-adult, lab-stock N2s and found them well fit with a 1% linear growth per hour: $R^2 = 0.92$ and 0.76, respectively. This may explain the observed hourly differences among animals. Likewise, we reasoned that although we control our lab to maintain it at 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

doi:10.1038/nmeth.2560

parameters, whereas the month of the year had a slight but significant effect on all parameters except coiling (**Supplementary Fig. 3**). So that seasonal effects were accounted for, all data were controlled by wild types collected within a 2-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 2 s.d. of mean difference. Moreover, 20 worms discriminate a single s.d. at over 80% power (**Supplementary Fig. 3**). For example, when testing forward speed, 20 worms discriminate a mean difference of at least 53 µm/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 ("Midbody Speed"), 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 showed some apparent differences; further testing, with subdivided feature measurements, uncovered related significant measures (available online in our database). Nonetheless, many features appeared 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 ($P = 2 \times 10^{-12}$), results correlating with the bordering phenotype of LSJ1. Additionally, LSJ1 was slightly (0.1 mm) but significantly shorter in length ($P = 2 \times 10^{-13}$) than N2. In comparison, the summary statistics for foraging are nearly overlapping ($P = 1$ for both sides) with a dorsal mean and s.e.m. of 37° ± 0.5° and 36.9° ± 0.4° for the LSJ1 and lab N2, respectively, and ventral values of 36.5° ± 0.3° for both strains.

Feature normalization. Each strain's measures were collapsed to the mean of their observations and normalized to their wild-type controls (through subtraction of the control mean and scaling by its variance). In four strains, measures were exclusively found in either 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 (v.1.6), standard Matlab toolboxes (v.2010a) or standard R packages (v.2.15) or were created as needed. Several publicly available MathWorks packages significantly facilitated the analysis and figures within this paper: the videoIO toolbox by G. Dalley, swtest function by A.B. Saïda, fexact function by M. Boedigheimer, export_fig function by O. Woodford, notBoxPlot function by R. Campbell and rdir function by G. Brown.

Worm Tracker 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 in the camera's 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 used a 640 × 480–resolution, 30-f.p.s. camera with a magnification that results in 3.5–4.5 µm per pixel and an FOV of roughly 2.5 × 2 mm^2 at a focal distance of nearly 1 cm. Older videos have a frame rate of 20–30 f.p.s. New videos maintain 30 f.p.s. 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 627 nm to avoid exciting a short-wavelength avoidance response through the *C. elegans* LITE-1 receptor. Finally, the motorized stage uses Zaber T-NA08A50 linear actuators. The stage travels 5 cm at up to 8 mm/s in orthogonal x and y axes with a resolution of just under 0.05 µm.

Worms were maintained as previously described in ref. 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 h before use. Strains are maintained on standard NGM plates[24] seeded with three drops of OP50. Six adult worms are transferred to a new plate to maintain stocks. At least two generations are passaged in these conditions before tracking. The evening before tracking, at roughly 5 p.m., L4 hermaphrodites are picked to a fresh plate, with ten worms per plate. The next morning, 3.5-cm low-peptone NGM plates are seeded with 20 µl of OP50, in the center of the plate, and allowed to dry. This 20-µl drop of OP50 is nearly circular and roughly 8 mm in diameter. The L4 worms from the evening before, now young adults, are transferred to the center

of the food on the 3.5-cm plates, one worm to a plate. The worms are given 30 min to habituate and are then tracked for 15 min. A wild-type N2 is always tracked, at the same time, on one of the eight nearby trackers to serve as a control. Strains are tracked between 8 a.m. and 6 p.m. across several hours and days and are randomly assigned to any of eight available trackers. In total (before filtering the videos for quality), approximately 25 worms are tracked per experimental strain, controlled by roughly 65 N2s tracked within 1 week of the experiments.

Worm segmentation. Video frames were extracted using the Matlab videoIO toolbox by G. Dalley. There is a sharp contrast between the worm and background in our video images. Worm pixels were segmented from the background using the Otsu method[25] to find a threshold. The largest eight-connected component in the thresholded image was assumed to be the worm. Frames in which the worm touched the image boundaries, was too small, lacked a clear head and tail or had unrealistic body proportions were not analyzed further. Frames containing stage movement were also removed to eliminate bad segmentations wherein the worm image may be blurred (see below: "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 erred on the side of caution and attempted to reject ambiguous segmentations rather than include them.

Once the worm was thresholded, its contour was extracted by tracing the worm's perimeter. The head and tail were located as sharp, convex angles on either side of the contour. The skeleton was extracted by tracing the midline of the contour from head to tail. During this process, widths and angles were measured at each skeleton point to be used later for feature computation. At each skeleton point, the width was measured as the distance between opposing contour points that determine the skeleton midline. Similarly, each skeleton point served as a vertex to a bend and was assigned the supplementary angle to this bend (**Supplementary Fig. 1b**). 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 used chain-code lengths[26]. Each laterally connected pixel was counted as 1. Each diagonally connected pixel was counted as $\sqrt{2}$. The supplementary angle was determined per skeleton point using edges 1/12 the skeleton's chain-code length in opposing directions along the skeleton. As there are an insufficient number of skeleton points to determine the angles at the ends of the skeleton, these angles were undefined (i.e., the first and last 1/12th of the skeleton had no bending angle defined). One-twelfth of the skeleton (**Supplementary Fig. 1a** 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 body-wall muscles and their innervation and cuticular rigidity[6].

Ventral side annotation and head detection. Before assigning the head and tail, videos were split into chunks in which worm skeletons could be confidently oriented with respect to each other. Chunk boundaries were set whenever there was a gap in skeletonized frames of 0.25 s or more. During these gaps, worm motion could make skeleton orientation unreliable. The skeletons within each chunk were aligned by determining which of the two possible head-tail orientations minimized the distance between corresponding skeleton points in subsequent frames. When possible, we unified chunks and healed up to 0.5-s interruptions by determining whether the worm was bent enough to achieve an omega turn and flip its orientation. If so, we traced the worm's path through its large bend to determine the new orientation. If the path could not be confidently traced, we avoided healing and maintained separate chunks.

The head was detected in each chunk of oriented frames. The head and neck perform more lateral motion (for example, foraging) even in uncoordinated mutants. Therefore, we measured lateral motion at both worm end points, across each chunk—unless the chunk was shorter than one-sixth of 1 s, which is too short to reliably measure such motion. In our setup, the head allows more light through than the tail and therefore appears lighter (has higher mean intensity). Therefore, we also measured the grayscale intensity at both worm end points, across each chunk. Linear discriminant analysis (LDA) was used on a combination of lateral motion and intensity at the worm end points 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 FOV. Prior to recording, all trackers were regularly calibrated to determine the conversion from pixels to absolute coordinates. When recording was complete, stage movements were matched to their video signature to convert segmented worms to absolute coordinates (offset by the stage's location).

During recording, every stage movement was logged. When recording was completed, the video was scanned to locate motion frames. Because recentering the worm causes an abrupt change in both the image background and the worm's location, these changes were simply measured as the pixel variance in the difference between subsequent frames. The Otsu method was used to find an appropriate threshold for delineating stage-movement frames. The number of stage movements and the intervals between them were matched against the log of software-issued stage-movement commands. If the match failed (an infrequent event usually caused by worms reaching the boundary of their plate or external factors damaging the recording), the worm and its video were discarded. In our data set, roughly 4% of the videos were discarded because of stage-movement failures.

With the stage movements matched to their video signature, the Otsu threshold was used once again to compute a local threshold that delineates a more accurate start and end for each individual stage movement. The same algorithm was also used 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 were discarded and each interval between stage movements was assigned a stage location. Thereafter, each segmented worm was converted to its absolute coordinates on the plate.

doi:10.1038/nmeth.2560

Feature overview. All feature formulas were computed from the worm's segmented contour and skeleton. The skeleton and each side of the contour were scaled down to 49 points for feature computation. Wild-type worms have four quadrants of longitudinal, staggered body-wall muscles[27]. 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, because the worm is confined to two dimensions, its body-wall muscles present roughly 24 degrees of freedom (although in practice it seems to be far less[8]). With 49 points we had two samples per degree of freedom and, therefore, expected to sample above the Nyquist rate for worm posture.

A common notation was used to define the body parts (**Supplementary Fig. 1a**). The head is controlled by the first four body-wall muscles per quadrant—approximately one-sixth the length of the worm[28]. Similarly, the neck is controlled by the next four body-wall muscles per quadrant—also approximately one-sixth the length of the worm. For this reason, we defined the head as the first one-sixth of the worm and the neck as the next one-sixth of the worm (skeleton points 1–8 and 9–16, respectively). For symmetry, we defined the tail and 'hips' in a similar manner, on the opposite end of the worm. The tail was the last one-sixth of the worm and the hips were defined as the next one-sixth (skeleton points 42–49 and 34–41, respectively). The midbody was defined as the remaining middle one-third of the worm (skeleton points 17–33). For some features, the head and tail were 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 were represented by top-level histograms and statistics as well as subdivisions exploring their values during forward, backward and paused states. This was to measure behaviors that depend on the state of motion, such as foraging amplitude, which is reduced during reversals in wild-type worms[29]. Many features were signed to reflect dorsal-ventral orientation, forward-backward trajectory and other special cases (for example, eigenworm projection) to capture any asymmetry. Finally, event-style features (coiling, turning, and motion states) were 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. See the **Supplementary Note** for further details.

19. Yemini, E., Kerr, R.A. & Schafer, W.R. *Cold Spring Harb. Protoc.* **2011**, 1475–1479 (2011).
20. Tsai, C.A. & Chen, J.J. *Bioinformatics* **25**, 897–903 (2009).
21. Zhao, B., Khare, P., Feldman, L. & Dent, J.A. *J. Neurosci.* **23**, 5319–5328 (2003).
22. McGrath, P.T. *et al. Neuron* **61**, 692–699 (2009).
23. Weber, K.P. *et al. PLoS ONE* **5**, e13922 (2010).
24. Brenner, S. *Genetics* **77**, 71–94 (1974).
25. Otsu, N. *IEEE Trans. Syst. Man. Cyber.* **9**, 62–66 (1979).
26. Freeman, H. *IRE Trans. Electron. Comput.* **EC-10**, 260–268 (1961).
27. Sulston, J.E. & Horvitz, H.R. *Dev. Biol.* **56**, 110–156 (1977).
28. White, J.G., Southgate, E., Thomson, J.N. & Brenner, S. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
29. Alkema, M.J., Hunter-Ensor, M., Ringstad, N. & Horvitz, H.R. *Neuron* **46**, 247–260 (2005).