Visible phenotypes based on locomotion and posture have played a critical role in understanding the molecular basis of behavior and development in Caenorhabditis elegans and other model organisms. However, it is not known whether these human-defined features capture the most important aspects of behavior for phenotypic comparison or whether they are sufficient to discover new behaviors. Here we show that four basic shapes, or eigenworms, previously described for wild-type worms, also capture mutant shapes, and that this representation can be used to build a dictionary of repetitive behavioral motifs in an unbiased way. By measuring the distance between each individual’s behavior and the elements in the motif dictionary, we create a fingerprint that can be used to compare mutants to wild type and to each other. This analysis has revealed phenotypes not previously detected by real-time observation and has allowed clustering of mutants into related groups. Behavioral motifs provide a compact and intuitive representation of behavioral phenotypes.

Results and Discussion

Using eight inexpensive USB microscope-based trackers we recorded high-resolution movie data of freely crawling worms covering 307 different mutant strains with a total of 7,708 individual worms. Worms were transferred to the surface of agar plates seeded with Escherichia coli OP50 and were allowed to habituate for 30 min before being recorded for 15 min at 25 frames per second. To automatically extract worm posture, the outline of the worm was determined after thresholding, and the skeleton was found by tracing the midline connecting the two points of highest curvature on the outline, which correspond to the worm’s head and tail. The head position was determined automatically using the distribution of pixel brightness and the magnitude of lateral motion of the head and tail (Fig. L4). Skeleton coordinates were converted to a position- and orientation-independent representation by taking 48 tangent angles evenly distributed from head to tail and subtracting off the mean angle (24).

It has been shown that the space of shapes explored by Caenorhabditis elegans during spontaneous behavior on agar without bacterial food is only four-dimensional (24). In other words, just four fundamental shapes, or eigenworms, can be added together in different proportions to reconstruct any worm posture. Because the four eigenworms (Fig. 1B; Fig. S1) provide an essentially complete description of worm posture, each frame in a movie of worm behavior can be represented as just four numbers, the amplitudes along each dimension when the shape is projected onto the eigenworms (Fig. 1C).

To use this representation as a common basis in behavioral genetics experiments, it must also capture mutant worm shapes, even though many mutants adopt postures that appear very different from those of wild-type animals. Even in wild type, locomotion is known to depend on environment; in particular, the presence of a bacterial food lawn significantly affects many aspects of locomotion (25, 26). Nonetheless, using data from wild-type (N2) worms tracked on food, we found that four eigenworms were sufficient to capture 93 + 3% (mean ± SD) of the variance of worm shapes (Fig. 1D). Likewise, when we projected behavioral data from 307 mutant strains (n = 7,708) onto the wild-type–derived eigenworms and used the four amplitudes to reconstruct the skeleton angles, the fit of the mutant data was comparable to wild type—92 ± 4% of the variance, including mutant shapes, is captured by the wild-type eigenworms (Fig. 1D). Even eigenworms derived directly from mutant data (Fig. 1B) show a remarkable reproducibility and similarity to the wild-type–derived eigenworms.

The fact that the wild-type–derived eigenworms also capture mutant postures may reflect fundamental constraints on worm behavior, with even highly uncoordinated mutants exploring different regions of essentially the same shape space. It is not clear what constrains worm behavior, but the mutant strains that are least-well fit by the wild-type eigenworms suggest some possibilities (Fig. S2). Among the worst-fit mutants are lon-2 (e67) which is longer than wild type, suggesting a role for body mechanics, and unc-4(gk705) and unc-34(e566), which affect synaptic specificity (27, 28), suggesting that neural circuit architecture also plays a role. Though these mutants may give insight into behavioral constraints, they are not alone sufficient to

A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion

André E. X. Brown, Eviatar I. Yemini, Laura J. Grundy, Tadas Jucikas, and William R. Schafer

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

Edited by Cynthia Kenyon, University of California, San Francisco, CA, and approved November 16, 2012 (received for review July 6, 2012)

Visible phenotypes based on locomotion and posture have played a critical role in understanding the molecular basis of behavior and development in Caenorhabditis elegans and other model organisms. However, it is not known whether these human-defined features capture the most important aspects of behavior for phenotypic comparison or whether they are sufficient to discover new behaviors. Here we show that four basic shapes, or eigenworms, previously described for wild-type worms, also capture mutant shapes, and that this representation can be used to build a dictionary of repetitive behavioral motifs in an unbiased way. By measuring the distance between each individual’s behavior and the elements in the motif dictionary, we create a fingerprint that can be used to compare mutants to wild type and to each other. This analysis has revealed phenotypes not previously detected by real-time observation and has allowed clustering of mutants into related groups. Behavioral motifs provide a compact and intuitive representation of behavioral phenotypes.

The study of unconstrained spontaneous behavior is the core of ethology, and it has also made significant contributions to behavioral genetics in model organisms. A powerful approach has been the careful expert observation of mutants to identify those with visible locomotor phenotypes, as demonstrated for many model organisms (1–6). However, as with most manually scored experiments, subjectivity can reduce reproducibility, whereas subtle quantitative changes or those that happen on very short or long time-scales are likely to be missed. Furthermore, manual observations are not scalable, and this has led to a widening gap between our ability to sequence and manipulate genomes and our ability to assess the effects of genetic variation and mutation on behavior.

Several recent reports describe systems that begin to address this gap by automatically recording and quantifying spontaneous behavior as possible. To do this, we use the inherent structure of a data set to identify informative behavioral parameters that may not be optimal for phenotypic comparisons and precludes the discovery of new behaviors that have not already been observed by eye. An alternative approach is to use unsupervised learning, which attempts to use the inherent structure of a data set to identify informative patterns; to do this, we first needed to extract worm postures from movie data and have as compact and complete a representation of worm behavior as possible.

Results and Discussion

Using eight inexpensive USB microscope-based trackers we recorded high-resolution movie data of freely crawling worms covering 307 different mutant strains with a total of 7,708 individual worms. Worms were transferred to the surface of agar plates seeded with Escherichia coli OP50 and were allowed to habituate for 30 min before being recorded for 15 min at 25 frames per second. To automatically extract worm posture, the outline of the worm was determined after thresholding, and the skeleton was found by tracing the midline connecting the two points of highest curvature on the outline, which correspond to the worm’s head and tail. The head position was determined automatically using the distribution of pixel brightness and the magnitude of lateral motion of the head and tail (Fig. L4). Skeleton coordinates were converted to a position- and orientation-independent representation by taking 48 tangent angles evenly distributed from head to tail and subtracting off the mean angle (24).

It has been shown that the space of shapes explored by Caenorhabditis elegans during spontaneous behavior on agar without bacterial food is only four-dimensional (24). In other words, just four fundamental shapes, or eigenworms, can be added together in different proportions to reconstruct any worm posture. Because the four eigenworms (Fig. 1B; Fig. S1) provide an essentially complete description of worm posture, each frame in a movie of worm behavior can be represented as just four numbers, the amplitudes along each dimension when the shape is projected onto the eigenworms (Fig. 1C).

To use this representation as a common basis in behavioral genetics experiments, it must also capture mutant worm shapes, even though many mutants adopt postures that appear very different from those of wild-type animals. Even in wild type, locomotion is known to depend on environment; in particular, the presence of a bacterial food lawn significantly affects many aspects of locomotion (25, 26). Nonetheless, using data from wild-type (N2) worms tracked on food, we found that four eigenworms were sufficient to capture 93 ± 3% (mean ± SD) of the variance of worm shapes (Fig. 1D). Likewise, when we projected behavioral data from 307 mutant strains (n = 7,708) onto the wild-type–derived eigenworms and used the four amplitudes to reconstruct the skeleton angles, the fit of the mutant data was comparable to wild type—92 ± 4% of the variance, including mutant shapes, is captured by the wild-type eigenworms (Fig. 1D). Even eigenworms derived directly from mutant data (Fig. 1B) show a remarkable reproducibility and similarity to the wild-type–derived eigenworms.

The fact that the wild-type–derived eigenworms also capture mutant postures may reflect fundamental constraints on worm behavior, with even highly uncoordinated mutants exploring different regions of essentially the same shape space. It is not clear what constrains worm behavior, but the mutant strains that are least-well fit by the wild-type eigenworms suggest some possibilities (Fig. S2). Among the worst-fit mutants are lon-2 (e67) which is longer than wild type, suggesting a role for body mechanics, and unc-4(gk705) and unc-34(e566), which affect synaptic specificity (27, 28), suggesting that neural circuit architecture also plays a role. Though these mutants may give insight into behavioral constraints, they are not alone sufficient to


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. E-mail: wschafer@mrc-lmb.cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211447110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1211447110

PNAS | January 8, 2013 | vol. 110 | no. 2 | 791–796
escape them completely: 86 ± 6% of the variance of un-2(ok3189) shapes (the worst-fit mutant in terms of rmsd) is still captured by the wild-type eigenworms.

Having established a compact and common basis for representing worm shapes, we next turn to the problem of automatic behavior detection. To avoid subjectivity, we used what might be considered a minimal definition of stereotyped behavior and simply searched for the most repetitive subsequence of a given length in a movie (Fig. 2A); in the amplitude representation, this is equivalent to time series motif-finding (29), and so we call these subsequences behavioral motifs. We used the minimum rmsd to define the best match and used either a brute force search or the MK algorithm (30) to find exact motifs.

We constructed a dictionary of 2,223 behavioral motifs by searching for motifs in movies sampled across all mutant strains (SI Methods). Each motif is a behavior performed by a single individual at two different times. The dictionary contains a wide range of behaviors, including short motifs essentially consisting of a single posture (Fig. 2B, motifs 1–3) as well as long and sometimes seemingly irregular behaviors that are nonetheless almost perfectly repeated (Fig. 2B, motif 14). Bouts of forward locomotion are relatively common but can have widely different frequencies and amplitudes (Fig. 2B, motifs 11–13). Finally, some motifs contain subtle behaviors that may not be picked in a manual inspection for stereotyped behaviors such as motif 9, which contains a pause interrupted by a small head bend.

We next used the motif dictionary to construct a quantitative phenotypic fingerprint for each recording. The fingerprint for each individual is a vector of distances between that worm’s behavior and each dictionary element (illustrated for three strains and 14 motifs in Fig. 2C). The distance between the fingerprints averaged over a strain provides a quantitative measure of their phenotypic dissimilarity. More specifically, for each element in the motif dictionary, we identified the best matching subsequence in the movie of interest and used the rmsd as the distance between a motif and a time series (31). To compare hundreds of strains, we found the distance between all of the motifs in the dictionary and each of the movies in the database. We then used a minimum redundancy maximum relevance (mRMR) criterion (32) to select 700 motifs for clustering. Reducing the number of motifs saves computational time, but the method is robust to exactly how features are selected: a random selection of 700 motifs also performs well. The results do not depend sensitively on the number of motifs. Qualitatively similar results are found using 300 motifs. Affinity propagation (33) was used to cluster mutants into phenotypically related groups using inverse Mahalanobis distance between strains as the similarity measure. We then resampled individuals from each group with replacement, recalculated the distance, repeated the clustering 100 times, and determined the frequency with which strains were in the same cluster. We kept only the most frequent cluster for each strain and illustrated these as edges in a phenotypic similarity network (Fig. 3; see Fig. S3 for a version with node labels).

The nodes in Fig. 3 are colored by phenotypic or molecular class. For a complete list of the strains and their corresponding class designation, see Dataset S1. There are four broad groups of mutants in the network: (i) on the lower left there is a cluster primarily of monoamine related genes (e.g., receptors, putative monoamine transporters, and enzymes involved in monoamine synthesis) and some transient receptor potential (TRP) ion channels; (ii) on the lower right there is a cluster of 64 G protein-coupled receptors; (iii) on the upper left is a loosely connected uncoordinated cluster; and (iv) on the upper right is a cluster of strains with nearly normal locomotion, including several mutant groups and the wild-type N2 strain itself. For the last group, it is important to emphasize that some of the mutants in the cluster with N2 are still significantly different from N2, as we discuss in more detail below.

To assess the clustering outcome in more detail, we compared the proximity of strains with predicted associations at increasing scales: loss-of-function alleles of the same gene, mutations affecting different genes but disrupting the same molecular complex, and at the largest scale, genes affecting a common molecular pathway. There were several loss-of-function deletion, nonsense, or splice-site mutant alleles of the same gene in the network, and these clustered significantly closer to each other than the network average (gene pairs are indicated by dashed red lines in Fig. 3): they were separated by an average of 1.6 edges, significantly less than the average network distance of 3.7 (P = 0.6 × 10−5, Wilcoxon rank-sum test). These genes, with the corresponding network distances are egl-21(n476) and egl-21(n6111) (one edges); ocr-4(tm2173) and ocr-4(s137) (two edges); trp-2; (gk298) (two edges);unc-10(m11177) and unc-10(e102) (two edges); unc-89(e1460) and unc-89(st85) (one edge); and trpa-2(ok3189), trpa-2(tm3085), and trpa-2(tm3092) (the ok3189 allele is two edges from the other alleles which are one edge from each other)}
other). The unc-108(n501) and unc-108(n777) (one edge) alleles are both dominant activating mutations in a Rab small GTPase (34) and are therefore also expected to result in a similar phenotype. Although there are two loss-of-function alleles of unc-4, these were not included in the analysis because one of them (gk705) is not part of the main network. There are also two pairs of genes that form subunits of the same complex: unc-38 and unc-63 encode subunits of the same acetylcholine receptor (35, 36), and unc-79 and unc-80 encode subunits of the NALCN neuronal sodium leak channel (37, 38). Both pairs of mutants cluster together in the network as expected (Fig. 3, Upper Left). syg-1 is required for specifying synaptic specificity and acts as a receptor for syg-2; however, they are separated by three edges in the network, just under the network average of 3.7.

We performed a similar analysis for genes in common molecular pathways. Fig. 4 highlights four examples of monoamine signaling (see Fig. S4 for acetylcholine receptors and pathways regulating synaptic release, insulin signaling, Go/Gq signaling, and mechanosensation). Most of the monoamine mutants are expected to have relatively subtle behavioral phenotypes, but they still form significantly tighter clusters than the network overall. The exception appears to be for serotonin; however, the outlying pair of genes cat-4 and bas-1 encode molecules required for both serotonin and dopamine biosynthesis—indeed, they cluster tightly with other dopamine-related genes. If we consider only the genes involved exclusively in serotonin signaling, we find that these too are significantly more tightly clustered than the overall network (Fig. 4).

To determine how the different components of the algorithm contribute to this clustering result, we repeated the analysis with different versions of some subroutines. In each case, nothing was changed about the analysis procedure except the specified substitution. Instead of the inverse Mahalanobis distance we used the inverse Euclidean distance between strains as the phenotypic similarity measure and found that the resulting network (Fig. S5) showed fewer of the predicted associations discussed above.
and had lower modularity (proportional to the number of within-
group connections minus the number of connections expected
for a random network of the same degree) (39). However, when
we used 700 randomly selected motives from the 2,223 element
dictionary instead of mRMR and repeated the clustering, the
resulting phenotypic network was of similar quality to the origi-

nal in terms of its modularity and the significant association
of molecular pathway components, although it has slightly fewer
expected allelic and molecular complex connections (Fig. S5B).
When short behaviors with the same lengths as the motives in
the dictionary are selected at random and used for clustering, we find
a similar-quality network with three fewer expected associations
(Fig. S5C). This finding suggests that selecting a large-enough
number of behavioral sequences is almost as good for phenotypic
profiling, and that stereotyped behaviors, at least according to
our minimal definition, are not specifically required to usefully
compare mutants.

In addition to identifying broad categories and analyzing path-
ways, we want to generate specific hypotheses about functional
interactions based on phenotypic similarity. To demonstrate the
potential of this approach, we considered two degenerin/epithelial
Na\(^+\) (DEG/ENaC) channels present in the N2-like cluster: asic-2
(ok289) (n = 19) and acd-5(ok2657) (n = 20). Neither of these
genes has a known function, nor do the deletion strains have a
previously reported phenotype. We used the mRMR criterion to
find the two most-distinguishing behavioral motives with respect
to N2 (Fig. 5). Qualitatively, in both cases we find one motif that
represents a bout of forward locomotion and another that is
a pause in a curved shape. The DEG/ENaC channel mutants are,
on average, further from the forward locomotion and closer to
the curved pause. Both differences are statistically significant
based on a Hotelling T\(^2\) test with permutation (40, 41) [asic-2
(ok289) vs. N2, \(P = 0.0019\); acd-5(ok2657) vs. N2, \(P = 9 \times 10^{-6}\)].
Comparing the two DEG/ENaC mutants to each other (Fig. 5C),
two motives are selected but there is no significant difference
between their distances to the motives (\(P = 0.796\), Hotelling T\(^2\)
with permutation). In other words, these two mutants were found
to be different from N2 yet they were not distinguishable from
each other using the same procedure. It should be noted that this
is not true for all mutants in the N2-like group; for example, two

![Phenotypic association network](image)

**Fig. 3.** Phenotypic association network. Nodes are mutant strains, and edges show phenotypic connections. Edge transparency indicates the fre-
quency with which two strains cluster together after resampling from the
data with replacement (frequently clustering strains are connected by dark
dges). The network layout is determined using spring embedding with
edge weights determined by the inverse phenotypic distance. Color-coding
indicates either known phenotypic classes or molecular pathway families.
*(Inset)* Network around N2 with increased transparency and smaller node
labels for clarity. The DEG/ENaC mutants discussed in Fig. 5 are shown with
a red rectangle.

![Molecular pathway networks](image)

**Fig. 4.** Genes involved in monoamine pathways cluster together. In each
panel, genes in the indicated class of monoamine signaling are highlighted in red. The mean ± SE of the shortest path connecting each pathway
member is listed below the network. Cases where the intragroup distance is
significantly smaller than the network overall based on a Wilcoxon rank-sum

test are highlighted in red. In the case of serotonin, the results are also
shown without cat-4 and bas-1 because they encode molecules required for
both serotonin and dopamine biosynthesis. Included genes are as follows:
dopamine and receptors: cat-4(e1112), dop-1(vs101), dop-1(vs100); dop-2
(vs105), dop-1(vs100); dop-1(vs105), dop-2(vs105), dop-1(vs100); dop-3
(vs106), dop-2(vs105), dop-1(vs100); dop-3(vs106), dop-3(vs100), dop-3
(tm1392), bas-1(ad446), cat-4(e1141). Serotonin and receptors: bas-1(ad446),
cat-4(e1141), ser-1(ok345), ser-4(ok512), ser-7(tm1325), tph-1
(mean shortest path: 1 ± 0
p = 2 × 10^(-2)

**Tyramine and Receptors**

(mean shortest path: 3.2 ± 0.4
p = 0.2 (without cat-4 and bas-1),
1.8 ± 0.2, p = 2 × 10^(-3))

**Octopamine and Receptors**

(mean shortest path: 2 ± 0.4
p = 0.009

(mean shortest path: 1.5 ± 0.2
p = 9 × 10^(-4))

**Dopamine and Receptors**

(mean shortest path: 1 ± 0
p = 2 x 10^(-2)

(mean shortest path: 3.2 ± 0.4
p = 0.2 (without cat-4 and bas-1),
1.8 ± 0.2, p = 2 × 10^(-3))

**Serotonin and Receptors**

(mean shortest path: 2 ± 0.4
p = 0.009

(mean shortest path: 1.5 ± 0.2
p = 9 × 10^(-4))

(mean shortest path: 1 ± 0
p = 2 x 10^(-2)

other DEG/ENaC channels near each other in the same cluster are still distinguishable using the same approach (Fig. S6). Given that DEG/ENaC channels are known to form heteromeric complexes (42), we arrive at the hypothesis that these channels share a similar function and may even operate in the same channel complex in some cells.

Phenotypic profiling using automatically extracted behavioral motifs can reveal abnormal phenotypes in mutants that were not apparent from manual observation. Furthermore, phenotypic associations can sharpen hypotheses of gene function, especially when combined with other information, such as sequence similarity, and can therefore help guide functional experiments. Although the discovery phase is unsupervised, the end result is still an intuitive summary of a phenotype in terms of a small number of short behaviors. Because the assumptions underlying behavioral motif extraction are minimal, we expect our method to apply generally to many model organisms. In particular, because zebrafish larvae are of fixed length and generate body movements in two dimensions, little modification would be required to adapt the approach described here. Although a representation based on skeleton angles would not be optimal for Drosophila larvae, a representation based directly on outline curvature could make Drosophila larva locomotion and in vitro cell motility (43) amenable to unsupervised motif discovery as well. Motif-derived phenotypes are related to functional classes but are derived completely independently from other data; we therefore expect them to provide complementary information that may be usefully combined with proteomic and transcriptomic data in the future.

**Methods**

**Strains.** All strains were maintained at 22 °C as previously described (44). A complete strain list is included in Dataset S1.

**Tracking.** Single-worm tracking was performed as previously described (45, 46). Briefly, single worms were transferred to agar plates seeded with 20 μL of OP50 bacteria using a platinum wire and allowed to habituate for 30 min. Then each worm was tracked for 15 min at 25 frames per second using one of eight trackers consisting of a USB microscope mounted on a motorized stage. The camera was moved beneath the crawling worm, which was kept stationary to avoid mechanically stimulating the worm. Each tracker was controlled by a computer running home-built software written in Java (www.mrc-lmb.cam.ac.uk/wormtracker).

**Segmentation and Skeletonization.** All analysis software described below was written in MATLAB. Each movie frame was segmented using the Otsu threshold (47), and the worm was taken to be the largest connected component in the resulting image. The curvature of the outline of this connected component was determined, and the two points of highest curvature were taken to be the head and the tail. The skeleton was found by tracing the midline of the outline between these two points. The skeleton was divided into 49 equally spaced points, which were used to define 48 tangent angles. The mean of these angles was then subtracted.

**Eigenworm Representation.** Eigenworms were derived from at least 1 h of pooled crawling data for each strain as previously described (24). The skeleton angles in each frame were then projected onto the wild-type–derived eigenworms, and these amplitudes were used for all further behavior analysis.

**Motif Dictionary.** Motifs were discovered as previously described1 with the following modifications. Motifs were discovered using distances across all four eigenworm amplitudes simultaneously, ensuring an essentially one-to-one correspondence between a time-series motif and a segment of worm behavior. Furthermore, no normalization was performed on the candidate subsequences to preserve the amplitude offsets and magnitudes that are essential for maintaining the times-series behavior mapping. Nine different length motifs from 1.6 to 32 s (40–800 frames) were discovered in each of 1,542 movies (each 15 min long) sampled from all mutant strains, resulting in an initial motif index with 13,878 entries. Motifs were discovered on data downsampled by a factor of four. The 20% worst matching motifs of each length were dropped, and from the remaining 80%, the 20% of each length with the highest variance across all dimensions were kept, leaving 2,223 motifs in the pruned dictionary. Keeping high variance elements ensures

---

Fig. 5. Maximally distinguishing behavioral motifs. For each of the indicated comparisons (A–C) the two most-distinguishing behavioral motifs from the dictionary are found using mMR. The plots on the left show the z-normalized distance between the compared strains and the motif (mean ± SE). The motif amplitudes and the corresponding worm postures are shown in gray. The colored lines show the mean-matching motifs from each of the compared strains. For example, acd-5(ok2657) matches the first motif in A more closely than N2 on average, and this is visible in the amplitude overlay.

---

Brown et al.
that the pruned dictionary would not contain only pauses, which tend to be important distinguishing features. The full-resolution versions of each motif defines the feature vector for each individual worm. The feature matrix consisting of the full set of feature vectors was further reduced to 700 features using mMR (32) to find the 700 most-informative motifs. The resulting feature matrix was 2-normalized to prevent outlying motifs from dominating the dissimilarity measure and allow subtle but potentially important differences from common motifs to contribute. Pair-wise Mahalanobis distances were calculated between all of the strains. Because the number of features is significantly greater than the number of individuals in each group, the Mahalanobis distance is calculated using a shrinkage estimate of the covariance matrix (47) for each strain.

Clustering. Clustering was performed using affinity propagation (33) with the inverse Mahalanobis distance as the similarity measure and the median similarity as the clustering preference factor. Individual worms were then resampled with replacement from each strain, and the clustering was repeated 100 times. The clustering frequency was determined, and we kept only the most frequent 10% of connections and illustrated these as edges in a phenotypic similarity network. Cytoscape (48) was used to display the network. The nodes are arranged using a force-directed layout with inverse Mahalanobis distance as the edge weight so that more similar highly interconnected strains are nearer to each other (49).

ACKNOWLEDGMENTS. The authors acknowledge Eamonn Keogh, Abdullah Al Mueen, Hanchuan Peng, Kevin Murphy, Andrew Duchi, and Suchi Sarai for sharing code; Roland Schwarz and Greg Stephens for useful discussions and sharing code; and Robyn Brantly for critically reading the manuscript. Some strains were provided by National Institutes of Health (NIH) Chemical Genomics Center Office of Research Infrastructure Program P40 OD010440. Support was also provided by Medical Research Council Grant MC-A022-59891, NIH Grant DA012841, and a long-term fellowship from the Human Frontier Science Program (to A.E.K.B.).
Supporting Information

Brown et al. 10.1073/pnas.1211447110

SI Methods

Motif Analysis and Clustering in Detail. The four-channel time series of eigenworm-projected amplitudes initially contains small gaps when frames cannot be segmented because of stage motion blur, worm coiling, or because the worm outline is obscured in thick food. These gaps are linearly interpolated to produce uniformly sampled data in each channel. This data are then downsampled 1:4 to speed up motif finding.

Motifs are then discovered in every fifth movie in the entire downsampling dataset with the following lengths: 10, 15, 20, 30, 50, 75, 100, 150, and 200 frames (these are downsampled lengths, so with the original data taken at 25 frames per second the shortest motifs correspond to behaviors lasting 1.6 s). Each motif pair is saved in an index that contains the parent filename, the start index of the first motif, the start index of the second motif, the motif length, the distance between the motif pairs, and the sampling rate, which corresponds to 13,878 motifs in the initial index [7,708 (number of individuals) \div 5 (only every fifth was searched) \times 9 (number of motif lengths used)]. Of these, we drop the 20% with the highest distance between the two sequences in the motif pair for each length.

The reduced motif index is then used to get full-resolution motifs (no longer downsampled 1:4) from each parent file. From these 11,106 motifs, we take the 20% with the highest variance averaged over all four channels to select for motifs that show some motion (although many still contain pauses at various postures), leaving 2,223 motifs in the final dictionary. Each movie in the dataset is represented by a vector of distances from elements in the motif dictionary. For each element in the dictionary, its best match is found in the movie by sliding along the projected amplitudes from the movie and calculating the rmsd at each position. The minimum over all positions is taken as the distance between the motif and the movie’s time-series. A vector of 2,223 distances (or features) is thus calculated for every movie. This matrix is normalized by subtracting the population mean and dividing by the population SD and then further reduced by selecting the top 700 features using minimum redundancy maximum relevance (mRMR) (1).

To determine the phenotypic distance between each strain, we calculate the Mahalanobis distance between them. A shrinkage estimate of the covariance matrix is calculated from each ∼20 individual \times 700 feature matrix. These are then pooled for each strain comparison:

\[
S_{\text{pooled}} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2},
\]

where \(S_1\) and \(S_2\) are the estimated covariance matrices of the two data sets and \(n_1\) and \(n_2\) are the number of individuals in each group. The Mahalanobis distance is then

\[
d^2_{\text{maha}} = (\mathbf{x}_1 - \mathbf{x}_2)^T S^{-1}_{\text{pooled}} (\mathbf{x}_1 - \mathbf{x}_2),
\]

where \(\mathbf{x}_1\) and \(\mathbf{x}_2\) are the mean vectors of each group. We calculate each pairwise distance in this way and use the element-wise inverse of this distance matrix as the similarity measure for affinity propagation (2).

We resample from the feature matrix with replacement, recalculate the Mahalanobis distance matrix, and recluster using affinity propagation 100 times. Two strains are said to be connected if they clustered together (i.e., shared the same exemplar) frequently in the 100 rounds, where frequently means in the top 10% of the frequency distribution overall; this is the network of connections plotted in Fig. 3 and Fig. S3 using a force-directed layout in Cytoscape with spring strength determined by inverse Mahalanobis distance and edge transparency by cluster frequency.


Fig. S1. Eigenworms are eigenvectors of the covariance matrix of skeleton angles. (A) The covariance matrix of skeleton angles for ∼2 h of pooled worm-crawling data. The diagonal from the lower left to upper right of the matrix shows the variance at each skeleton point from head to tail. (B) The fraction of the skeleton variance that is captured increases as more eigenworms are used. After four eigenworms, the improvement is relatively small. (C) The first six wild-type (N2) eigenworms.

Fig. S2. A gallery of worm shapes illustrating the wide range of shapes that are adopted by the wild-type and three of the worst-fit mutants. The upper part shows frames from tracking movies with the outline and skeleton from the segmentation algorithm overlaid. The lower part shows the raw skeleton in orange with the reconstructed skeleton in gray. The reconstructed skeleton is generated by adding together four eigenworms in different proportions and integrating the tangent angles from head to tail. Images and skeletons have been rotated to have a mean skeleton angle of zero. The asterisk indicates a lon-2 worm that has five body bends; this is more than the wild type adopts and so the four wild-type-derived eigenworms cannot accurately fit all of the bends. The fraction of the variance captured by four eigenworms is shown below each group of images (mean ± SD).
Fig. S3. A high-resolution version of the cluster network shown in Fig. 3 that includes node labels.
**Total Network**

- **Path Length Distribution**
  - Mean shortest path: 3.66 ± 0.01

**Insulin Signalling**

- **Included Genes:**
  - ins-4(ok3534), ins-5(ok3297), ins-31(ok3543), ins-30(ok2343), ins-28(ok2722), ins-27(ok2474), ins-25(ok2773), ins-22(ok3616), ins-18(ok1672), ins-16(ok2919), ins-15(ok3444), ins-11(tm1053)
  - Mean shortest path: 2.3 ± 0.1
  - \( p = 3 \times 10^{-14} \)

**Acetylcholine Receptors**

- **Included Genes:**
  - acr-10(ok3064), acr-11(ok1345), acr-14(ok1155), acr-15(ok1214), acr-18(ok1285), acr-19(ad1674), acr-20(ok1887), acr-21(ok1314), acr-23(ok2804), acr-3(ok2049), acr-6(ok3177), acr-7(tm1863), acr-9(ok9133), gar-2(ok520), lev-11(e427), lev-8(e153), unc-29(e193), unc-38(e264), unc-62(ok1075)
  - Mean shortest path: 2.3 ± 0.1
  - \( p = 2 \times 10^{-21} \)

**Synaptic Release**

- **Included Genes:**
  - rab-3(y250), nic-19(ok833), unc-10(md1117), unc-10(e102), unc-2(gk366), unc-2(ax106), unc-26(m2), unc-35(e950))
  - Mean shortest path: 2.1 ± 0.1
  - \( p = 3 \times 10^{-5} \)

**Go/Gq Signalling**

- **Included Genes:**
  - goa-1(sa734), egl-30(ep271), egl-30(n686); goa-1(n1134), eat-16(sa609)
  - Mean shortest path: 3 ± 0.4
  - \( p = 0.9 \)

**Touch Neuron Function**

- **Included Genes:**
  - mec-10(e1515), mec-10(tm1552), mec-10(u20), mec-12(e1605), mec-12(u76), mec-14(u55), mec-14(u228), mec-4(u253), mec-7(u448)
  - Mean shortest path: 2.1 ± 0.1
  - \( p = 2 \times 10^{-9} \)

---

**Fig. S4.** Functional pathways cluster together. (Upper Left) Distribution of path lengths between all nodes in the network. All other panels show the positions of functionally related genes in the network. Most of the pathways cluster significantly more tightly than the network over all based on a Wilcoxon rank-sum test. The genes included in each pathway are listed beneath each network plot. Some gene lists have more entries than there are points indicated in the network because some of the genes are not present in the main network (Fig. 3).
Fig. S5. The phenotypic network is not sensitive to changes in some subroutines. The number of predicted associations that are present in the network as well as the network modularity (proportional to the number of connections within groups minus the number of such connections expected for a random network with the same degree) were used to evaluate clustering outcome. The groups used for the modularity calculation are those indicated by color in the network plots and listed in the strain list (Dataset S1). (A) The network shown in Fig. 3 constructed as described in the methods. (B) If 700 motif distances are chosen randomly from the 2,223 instead of using mRMR to select 700, the modularity does not decrease and the same pathways are significantly associated as in the original network. The number of expected allelic and molecular complex connections is somewhat reduced because trpa-2(tm3092) and one of the alleles of unc-89 are not part of the main network. (C) Selecting subsequences with random start positions instead of repeated subsequences has a similar effect as eliminating feature selection in terms of expected associations: it does not decrease the modularity, and the same pathways are significantly associated as those for the motif-based network. Again, the number of expected allelic and molecular complex connections is somewhat reduced because trpa-2(tm3092) and both alleles of unc-89 are not part of the main network. (D) Using Euclidean distance to measure the phenotypic distance between strains instead of Mahalanobis distance decreases network modularity, and only the acr, dop, and synaptic release pathways are significantly associated. Of the allele pairs and molecular complexes, only three are connected: trp-2 (three edges), egl-21 (two edges), and syg-1 and syg-2 (one edge).
**Fig. S6.** Maximally distinguishing behavioral motifs. For each of the indicated comparisons (A–C), the two most-distinguishing behavioral motifs from the dictionary are found using mRMR. The plots on the left show the z-normalized distance between the compared strains and the motif (mean ± SE). The motif amplitudes and the corresponding worm postures are shown in gray. The colored lines show the mean-matching motifs from each of the compared strains. Both degenerinepithelial Na+ channels are distinguishable, in contrast to *acd-5* and *asic-2*, as shown in Fig. 5.

**Other Supporting Information Files**

Dataset S1 (XLS)