Supplementary Table 1. List of strains. Every assayed strain is listed alongside its genotype and annotations. Strains with no previously characterized phenotype (as of Wormbase version 233, 07/10/2012) are highlighted.

Supplementary Table 2. List of all 702 features. Each feature is listed with its name, up to 2 subdivisions, and the units of measure. D = dorsal, V = ventral.

Supplementary Table 3. Phenotypic ontology and significance. Each strain is listed alongside its genotype and phenotypic ontology. “–” means the associated feature(s) measured less than the control, “+” means they measured greater, and “Δ” means they measured different (but the difference cannot be quantified as simply lesser or greater than). The minimum q-value is shown, per group, for the Wilcoxon rank-sum tests and the multivariate Hotelling T2.

Supplementary Table 4. Phenotypic ontology and significance for the case where outliers were removed (see Supplementary Note for details). Each strain is listed alongside its genotype and phenotypic ontology. “–” means the associated feature(s) measured less than the control, “+” means they measured greater, and “Δ” means they measured different (but the difference cannot be quantified as simply lesser or greater than). The minimum q-value is shown, per group, for the Wilcoxon rank-sum tests and the multivariate Hotelling T2.

Supplementary Table 5. Summary data for all strains and features. The first sheet of the table has mean feature values for each strain for all feature measures. The second sheet shows the q-values for the comparison of each strain to the lab N2 data.
Supplementary Figure 1. Feature computation. (a) The locations of the body parts used for feature computation: head, neck, midbody, hips, and tail. Each body part encompasses 1/6 of the worm (measured along the contour and skeleton, independently), save for the midbody which is 1/3. The head and tail are further split in half to create the head and tail tips, each 1/12 long. (b) The bend angle ($\alpha$) is the difference in tangent angles at each point; or, alternatively phrased, the supplementary angle ($\alpha$) with respect to the angle formed by any
three consecutive points ($\beta$). The bend angle is signed negatively whenever the ventral side is concave within the bend (as is the case for the bend shown). (c-f) Methods of quantifying worm motion. (c) A simple diagram represents worm velocity and the crawling wave. (d) The velocity vector of a body part is measured relative to the head-tail axis. The velocity, per body part, is the vector of its respective centroid. (e) A crawling wave is represented as the sinusoidal wave of the bend angle at its associated body part. Note the dorsal-ventral asymmetry both in the amplitude and in the wavelength itself. The dotted window encloses a waveform used to measure an instantaneous crawling wave. The Fourier transform of the waveform reveals a strong peak. This peak defines the instantaneous crawling amplitude and frequency. (f) A 25 second window of worm motion reveals 2 forward, 3 backward, and roughly 5 paused events. A short, small peak (between the second and third pauses) that may have been forward motion, remains unclassified due to ambiguity. (g) Worm foraging is measured from the bend angle between the two sections of the head (panel a). The noisy signal (black) is smoothed (red) by convolving with a Gaussian. The foraging amplitude is defined as the largest foraging bend angle measured, prior to crossing 0°. Foraging speed is simply the angular speed. Ventral and dorsal foraging is present within the first 1 second of the trace. The latter half of the trace displays the difficulties associated with measuring signal above the noise. At nearly 1 second in, the nose appears to quickly cross ventrally before rebounding dorsally. Smoothing eliminated the associated sign change in amplitude; although, a small, nearly 0° ventral amplitude may well be considered noise. (h) The range is defined, per frame, as the distance of the worm’s midbody from its final path centroid. The central dot displays the final path centroid. The two arrows display the range at early and late times within the experiment. (i) The locations of worm dwelling are shown as a heatmap. A single location of dwelling dominates faint traces of the worm’s path during motion.
Supplementary Figure 2. WT2 sensitivity. (a-c) The habituation and growth are shown for 25 young-adult, hermaphrodite N2s averaged in 1 minute bins. (a) The habituation of the crawling-wave amplitude is fit with an exponential with time constant \( \tau = 48 \) minutes (\( R^2 = 0.59 \)). (b) The habituation of the crawling-wave frequency is fit with an exponential with a time constant \( \tau = 19 \) minutes (\( R^2 = 0.71 \)). (c) Both length and width are fit with a 1% linear growth per hour (\( R^2 = 0.92 \) and 0.76 respectively). (d-e) The means are shown as vertical lines for positive (solid line) and negative (dashed line) portions of the histogram. Wilcoxon rank-sum tests were performed with Bonferroni correction for the 12 tests performed across both panels. (d) 21 N2s from CGC stock (green) are compared to 27 N2s from our lab stock (gray). The mean and SEM, for the CGC and lab respectively, are: length = 1162 ± 13\( \mu \)m, 1183 ± 19\( \mu \)m (\( p = 1 \)); forward speed = 137 ± 14\( \mu \)m/s, 104 ± 12\( \mu \)m/s (\( p = 0.5 \)); backward speed = 142 ± 15\( \mu \)m/s, 98 ± 11\( \mu \)m/s (\( p = 0.3 \)); dorsal foraging = 42 ± 0.6°, 38 ± 0.6° (\( p = 6 \times 10^{-3} ** \)); ventral foraging = 41 ± 0.8°, 39 ± 0.6° (\( p = 1 \)); exploratory range = 2.7 ± 0.3mm, 1.8 ± 0.2mm (\( p = 0.2 \)). (e) 43 LSJ1s (orange) are compared to 107 N2s from our lab stock (gray). The mean and SEM, for the LSJ1 and lab respectively, are: length = 1049 ± 9\( \mu \)m, 1166 ± 7\( \mu \)m (\( p = 2 \times 10^{-13} **** \)); forward speed = 213 ± 6\( \mu \)m/s, 106 ± 6\( \mu \)m/s (\( p = 6 \times 10^{-14} **** \)); backward speed = 221 ± 6\( \mu \)m/s, 97 ± 10\( \mu \)m/s (\( p = 2 \times 10^{-14} **** \)); dorsal foraging = 37 ± 0.5°, 36.9 ± 0.4° (\( p = 1 \)); ventral foraging = 36.5 ± 0.3°, 36.5 ± 0.3° (\( p = 1 \)); exploratory range = 3.3 ± 0.1mm, 2.1 ± 0.1mm (\( p = 2 \times 10^{-12} **** \)).
Supplementary Figure 3. Wild type variability. 1,218 young-adult N2 hermaphrodites were recorded over 3 years spontaneously behaving on food. (a) Our lab-stock wild type show significant differences in hourly and monthly measures at $p \leq 0.05$, but none daily. Kruskal-Wallis, one-way analysis of variance, tests were performed with Bonferroni correction for 18 tests of 3 groups (hour, day, and month) by six common measures: length ($p$-value hour = $4 \times 10^{-2} *$, day = 0.6, month = $4 \times 10^{-9} ****$), forward speed ($p$-value hour = $5 \times 10^{-8} ****$, day = 1, month = $1 \times 10^{-4} ***$), foraging amplitude ignoring the dorsal-ventral orientation ($p$-value hour = $2 \times 10^{-7} ****$, day = 0.5, month = $3 \times 10^{-8} ****$), reversal frequency ($p$-value hour = $8 \times 10^{-3} **$, day = 0.3, month = $1 \times 10^{-4} ***$), coiling frequency ($p$-value hour = 1, day = 1,
(p-value hour = $2 \times 10^{-23}$ ****, day = 0.3, month = $5 \times 10^{-14}$ ****). Each experimental mean is plotted as a black dot and the mean of means is shown in yellow, SEM in magenta, and standard deviation in cyan. (b) $\beta$ (the probability that a false null hypothesis will not be rejected, equal to 1 – statistical power) as a function of the number of worms observed in an experiment. $\beta$ is plotted for cases where the mean difference between the test case and the lab N2 is 2 (red), 1 (orange), 0.5 (green), 0.25 (cyan), and 0.125 (blue) standard deviations (SD) from the collective N2 data. Approximately, 10 worms discriminate a displacement of 2 SD with over 90% power (gray, dotted lines parallel to the axes) and 20 worms discriminate 1 SD at over 80% power (black, dashed lines parallel to the axes). Corresponding feature labels are presented on the far left, in panel a. Each plotted value was computed using 10,000 bootstrapped Wilcoxon rank-sum tests.

Since all tracked animals were identified as fourth-stage larvae the night before, the small but significant increase in length during the day is most likely simply due to age. The monthly variation may be due to differences in lab temperature or humidity, which can both change seasonally.
Length
Midbody Width
Bend Count
Curl Frequency
Forward Motion Time Ratio
Paused Motion Time Ratio
Backward Motion Time Ratio
Forward Midbody Speed
Backward Midbody Speed
Abs. Midbody Crawling Amplitude
Abs. Midbody Crawling Frequency
Abs. Foraging Amplitude
Abs. Foraging Speed
Omega Turn Frequency
Girth Range
Abs. Forward Path Curvature
Abs. Head Bend Mean
Abs. Midbody Bend Mean
Abs. Tail Bend Mean
Max Amplitude
Primary Wavelength
Eccentricity
Backward Motion Frequency
Abs. Backward Foraging Amplitude
Abs. Backward Foraging Speed
Abs. Backward Midbody Crawling Amplitude
Abs. Backward Midbody Crawling Frequency
Binned q-values
Supplementary Figure 4. 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. These data have not been clustered to facilitate finding a gene of interest.
Supplementary Figure 5. The clustering is shown for all 305 strains alongside the Schafer Lab N2. 213 features were eliminated due to poor statistical discrimination. Principal Components Analysis (PCA) was used to convert the remaining 489 features to eigenfeatures, discarding those accounting for the last 1% of variance. Hierarchical clustering was performed using uncentered correlation and complete linkage. Multiscale resampling was run...
with 10,000 bootstraps to generate approximately unbiased (AU) p values, measuring the confidence of tree selection for each branch. For easier viewing, the dendrogram has been divided into four sections arranged vertically. The arrangement of each part is indicated by boxes on the total dendrogram at the top of the page. The horizontal length of each branch in each partial dendrogram is scaled to its correlation value. The horizontal branches are color coded to show the AU of the clusters below them. AU $\geq 99.9\%$ is colored red, AU $\geq 99\%$ orange, AU $\geq 95\%$ yellow, and all else is black.
Supplementary Fig. 6. New locomotion phenotypes for three TRP channels. Both trpa-2 and ocr-4 have no characterized phenotypes despite previous experimental publications. We used Wilcoxon rank-sum tests and False-Discovery Rate (FDR) to determine q-values for each comparison. Corrections were performed for 329 groups (strains and time-based collections of N2s) by 702 features. Significance was set at q \leq 0.05. Each experimental mean is plotted as a black dot and the mean of means is shown in yellow, SEM in magenta, and standard deviation in cyan. For each allele, the experiments are brighter and shown on left, whereas the controls are darker and shown on right. (a) All three trpa-2 alleles (ok3189, tm3085, and tm3092; trpa-2 N = 19, 35, 38; control N = 41, 133, 93) show significantly more bends when reversing (q = 6x10^{-3}, 7x10^{-3}, 2x10^{-5}). In further support, all three alleles show a decrease in their second eigenworm projection when reversing (q = 7x10^{-3}, 2x10^{-3}, 4x10^{-5}).
measure of sinusoidal posture. (b) Both trp-2 alleles (gk298 and sy691; trp-2 N = 18, 21; control N = 49, 55) display significantly altered activity during reversals as well. Their waveform, when reversing, is more compact (q = 1x10^{-4}, 4x10^{-5}). Their foraging speed, when reversing, is faster (q = 4x10^{-6}, 9x10^{-3}). Interestingly, omega turns, which happen infrequently, also showed up as significant. The two trp-2 strains spend even less time in omega turns than their wild-type controls (q = 1x10^{-2}, 2x10^{-2}). (c) The ocr-4 alleles (tm2173 and vs137; ocr-4 N = 22, 22; control N = 55, 49) also display significant differences when subdividing by motion state. Both alleles move their tail more when paused (q = 9x10^{-5}, 3x10^{-3}). Furthermore, both ocr-4 alleles display smaller amplitudes when crawling (q = 2x10^{-3}, 5x10^{-3}).
Supplementary Note

Details of feature files and algorithms for feature measurement

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
described\(^8\).

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 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 1/30Hz (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 $\alpha$ 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 $0^\circ$. 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 upsilon turn events are computed similarly to a previously described method but using skeleton bends instead of a single head-midbody-tail angle. Omega and upsilon 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 $30^\circ$ at the first but not the last third of the body (the head but not the tail). The middle third must then exceed $30^\circ$. And finally, the last but not the first third of the body must exceed $30^\circ$ (the tail but not the head). This sequence of a $30^\circ$ 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 $15^\circ$ on one side of the worm (the first or last third of the body) while not exceeding $30^\circ$ 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 (Δ) 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 testing. 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 upsilon 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/](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 ± 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 equivalents\(^{10}\).
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 $q \leq 0.05$, 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) overlayed 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”
“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”).