UNIVERSITY OF CAMBRIDGE

An online database for behavioural phenotypes of the nematode *Caenorhabditis elegans*

Tadas Jucikas

Darwin College

A thesis submitted on the 15th of August, 2013 for the Degree of Doctor of Philosophy

I would like to dedicate this thesis to my loving grandmother Gabrielė, devoted parents Janina and Pranciškus and caring siblings Kristina and Viktoras. I am humbled and truly grateful for your support.

Abstract

Advances in the fields of neurobiology, genomics, computer imaging and bioinformatics provide a favorable landscape to study genetic and molecular mechanisms of model organism behaviour. However, manual animal observation and characterization has been an intrinsically subjective and time consuming effort. It encouraged the development of computer aided automatic tracking systems capable of recording animal behaviour and storing the data for further analysis.

For the nematode *C. elegans*, an array of worm trackers have been built focusing on high spatial and temporal resolution or on optimizing for higher throughput and imaging multiple worms at once. Early worm trackers quantified animal locomotion and described diverse groups of neural functions like proprioception, synaptic transmission and neuromodulation. However, behavioural phenotypes have been described only for a small fraction of genes despite evidence that the majority of genes are needed for wild type fitness. Emergence of reverse genetics and RNA interference based screens have greatly outstripped the phenotyping capability of *C. elegans* research community. In addition, the algorithms and techniques used were not based on community standards and prevented pooling behavioural data collected in different research groups.

To address the challenges of large scale behavioural phenotyping screens this thesis describes a platform that uses single worm trackers for high throughput analysis of behaviour experiments. The platform consists of an array of computers used for data collection and analysis, a MySQL database backbone and a set of MATLAB based tools for behaviour feature extraction and visualization. In addition to building a standalone high throughput analysis platform this work also proposes a standard for *C. elegans* neurobiology community and opens up a possibility of a centralized *C. elegans* behavioural experiment repository that would benefit from the work carried out by different research groups.

The thesis also presents an online *C. elegans* behaviour database that is inte-

grated with the data acquisition and analysis platform. The database hosts all of the behavioural experiment recordings and their analysis results. Users can access more than 10,000 experiments representing 329 strains, inspect their aggregate statistics and download the data for further analysis. Behavioural experiment videos can be viewed via the video sharing website YouTube and are embedded in the database pages. The online database address is http://wormbehavior.mrc-lmb.cam.ac.uk.

This thesis also illustrates how the database could be used to carry out novel behavioural feature extraction and analysis by accessing strain feature data. *C. elegans* navigation behaviour was studied in greater detail with a focus on turns, reversals and their related properties. One of the novel reorientation phenotypes that have been identified manifests in a loss of ventral bias for two *DEG/ENaC* channel mutants *asic-2 (ok289)* and *acd-5 (ok2657)*. This novel phenotype indicates functional similarity of these two genes. Similarly to that, other behaviours can be extracted and studied by downloading the feature data available via the online database.

The platform presented in this thesis enables quantitative behavioural analysis of *C. elegans* mutants and enables detection of previously uncharacterised locomotion phenotypes. It presents a dataset of more than 10,000 behaviour experiment recordings describing 329 *C. elegans* strains and is organized in a user friendly online database. It is to our knowledge the largest existing *C. elegans* behaviour dataset and it promises to be a rich resource for further exploratory analysis. The data and the methods presented in this thesis promote and inform further work in this area as well as encourage emergence of similar platforms in other model organisms.

Declaration

This thesis:

- is my own work and contains nothing which is the outcome of work done in collaboration with others, except where specified in the text;
- is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other university; and
- does not exceed the prescribed limit of 60,000 words.

Tadas Jucikas
August, 2013

Acknowledgements

The work for this thesis has been carried out at the Medical Research Council Laboratory of Molecular Biology (LMB), a research institute at the University of Cambridge. I would like to thank my supervisor Dr. William Schafer for giving me the opportunity to work on this challenging project. Dr. Schafer has given me a tremendous amount of freedom and maintained utmost faith in me over the years for which I am truly grateful. I was given an opportunity to present my work in a number of conferences, participate in collaborations and visit other laboratories. The skills and knowledge that I gained while working in Dr. Schafer's group will benefit me for many years to come. A big thanks also goes to my second supervisor Dr. Sarah Teichmann. I have never been turned down when contacting Dr. Teichmann for advice and have always been welcomed with enthusiasm. It was extremely important to know that I could always turn up and discuss a challenging bottleneck that I was facing.

I had a pleasure to work together with extremely motivating and inspiring individuals at Dr. Schafer's group. A special thanks goes to André Brown for discussing the progress of my work in great detail and always suggesting the most insightful observations about the methods that I was wrestling with or results that I tried to understand. I would also like to thank Buyun Zhao for helping me to grasp the beauty and complexity of *C. elegans* genetics as well as allowing me not to feel alone when turning up to the lab in a jacket and tie. A special thanks goes to Laura Grundy for painstakingly tracking thousands of animals and for being extremely understanding when changes and alterations were necessary. The patience that you displayed was colossal for which I am truly grateful. I thank Victoria Butler for tremendous amount of work that she did at the early stages of the project. The tracking protocol was defined with her help as well as the first recordings which allowed me to start tackling analysis algorithms. I would also like to thank Eviatar Yemini for his support and guidance when building the

analysis pipeline. The contribution of Eviatar was central to the completion of this work. I will continue to be motivated by the ambition, competitiveness and never ending energy displayed by him over the years of our work together.

The group would not have been as enjoyable without Robyn Branicky who excited everyone about important social activities and maintained the spirit of the group. Denise Walker has worked tirelessly to keep everything organized and put a huge amount of effort when the group had to move to the new LMB building. With her care and attention the tracker and computer equipment reached the new building without a single scratch. I would like to thank Marios Chatzigeorgiou for providing valuable advice while refining the pipeline. I also thank Barry Bentley for injecting fresh energy into the project and adding much needed attention to detail. I am very happy to leave the tracker system in his capable hands.

I thank former members of Dr. Schafer group Ithai Rabinowitch, James Cregg, Yosinori Tanizava and Marina Ezurra. Their support at the early days of the project was very important. Ithai and James especially have poured their energy into the project and allowed it to gain momentum much faster. I would also like to thank the neighbouring Dr. Mario De Bono group and its members for continuing support. I am also grateful to Dr. Tobias Klöpper for guiding me through the Internet and database technology and providing encouragement during the most challenging parts of the PhD.

My work at the LMB could not have progressed smoothly without extremely talented and caring support staff. I would like to thank the divisional administrators - Caroline Winsor, Kate French, Alison Turnock and Ami Cull as well as postgraduate managers Liz Pryke and Amie Blake for taking care of me. I would like to thank the LMB IT office and the scientific computing department, especially Jake Grimmett for taking care of all of my requests, as outrageous as they were at times.

In addition to everyone in the laboratory I have been extremely fortunate to spend time with remarkable individuals at the University of Cambridge. They have surrounded me in Fitzwilliam and Darwin Colleges as well as the societies that I had a pleasure to be part of. Some of the friends that provided me with much needed company were Sasan Behravesh, Progromakis Parasoglou, Aaron A. Bloch, Rory Stark, William Eucker, Zhenwei Peng, Vladimir Orlov, Sohan Dasgupta, Artūras Ratkus, Ignas Budvytis, Alex Tayler, Sebastian Pender, Jamie Blaza, Alex Hampton, Brent C. J. Clickard, Dwayne R. Menezes,

Tomoyuki Kuwaharada, Aaron Ralby, Simon Brunner and many others.

In addition to the friends in Cambridge I have received support from the ones further away as well. I would like to thank Arvydas Bložė, Andrew J. Knott, Peras Vaičius, Rasa Balsytė, Aistė Anusaitė-Daubaras and Darius Daubaras, Rytis Vitkauskas, Laura Dzelzytė, Jorge and Isabel Dorado and many others. I am truly grateful to Nikhil Sasidharan for his friendship and also for taking his time to proof read this thesis. I would also like to thank Mouna Chraibi for her support and faith in me during the most trying times.

The acknowledgements would not be complete without my teachers who supported and nourished my development. Only due to their continuous educational efforts, patience and confidence in my ability I was able to embark on this research project. I would like to thank the principal of my school in Lithuania, Bronislovas Burgis for opening up my mind and encouraging me to believe in my own abilities. I would also like to thank my Bachelor of Science thesis supervisors Dr. Martin Zacharias and Dr. Frank Oliver Glöckner for fascinating me with the field of Computational Biology, guiding me every step of the way and challenging me with the first truly difficult problems. I am extremely thankful to Dr. Roland Krause for hosting me at the Max Planck Institute for Infection Biology in Berlin for my first exciting internship that encouraged me to continue pursuing academic goals. Lastly, I would like to thank my Master of Philosophy thesis supervisor Dr. Pietro Lió for his trust and support.

Last but not least I am exceptionally grateful to my loving family for their tremendous devotion. My grandmother Gabrielė and parents Janina and Pranciškus have tirelessly strived to give me what they could not have and enabled me to follow my dreams. My brother Viktoras has worked hard to support me throughout my studies and I can not thank him enough for what he has done. My dear sister Kristina has also contributed enormously at every step of the way. She has shared the joys and hardships and has always poured endless energy and support to help me succeed. This thesis is dedicated to my family to honour the love that they have given.

Lastly, I would like to thank everyone else who I was not able to mention in this acknowledgement. I have been extremely fortunate to received support from many people that I have encountered during my education that spanned three countries and took over ten years. I am humbled and grateful for your support.

Contents

chapter 1

Introduction

Observing animals in a controlled environment as well as the wild has been a primary way to study their behaviour. With recent rapid advances in genetics, computer technology and statistical methods the concepts of Ethology can be further explored in a laboratory setting. By isolating behavioural mutants, cloning mutant genes and describing the neural and molecular basis of behaviour of a wide array of model organisms an unprecedented detail and scope of understanding can been reached.

In order to study complex mechanisms of behaviour a variety of genetic approaches are used. In forward genetic screens organisms can be isolated after a round of mutagenesis and selected on the basis of their distinct phenotypes. After phenotypic identification the genes can be mapped and underlying molecular mechanisms studied in greater detail. With the advent of automated DNA sequencing and the availability of whole genome sequences reverse genetics experiments became available. In contrast to forward genetics, reverse genetics involves selectively disabling known or predicted genes and looking for noticeable phenotypes. This allowed to deepen the scope of defining the genetic aspects of the model organism's nervous system.

Reverse genetics approaches rely on the ability to study resulting phenotypes in great detail. Phenotyping a large amount of organisms is a challenging and time consuming task. To overcome this bottleneck it is essential to make use of an array of computational biology and bioinformatics techniques. With the use of computers researchers can create high-throughput quantitative phenotyping

platforms to exhaustively study molecular aspects of behaviour.

The work presented in this thesis uses a nematode *Caenorhabditis elegans* and describes a quantitative phenotyping platform that allowed collecting the largest amount of nematode behavioural data to date. The data and insights gathered are presented in an online behavioural database and form an extremely rich resource for exploratory data analysis.

1.1 *C. elegans* - a powerful model organism

C. elegans is a free-living, transparent nematode that lives in temperate soil and bacteria rich environments. It's biology is simple yet it has many of the organ systems present in other animals. It has two sexes - hermaphrodites and males, and reproduces either by cross - or self - fertilisation. Therefore, mating behaviour is not necessary for maintaining mutant strains. The animal progresses through four stages of development (L1-L4) in approximately 4 days and can lay around 300 eggs. When crowded or in absence of food the nematode can enter alternative larval stage called the *dauer* state. Ecologically it uses this state to transfer from one place to another by attaching to insects or snails [1], [2]. The basic anatomy of the *C. elegans* includes a mouth, pharynx, intestine, gonad and collagenous cuticle. Reproduction organs in males consists of a tail specialized for mating, a single-lobed gonad and a *vas deferens*. Hermaphrodites have two ovaries, oviducts, spermatheca and a single uterus. The four bands of muscles running along the full length of the body are connected to neurons that allow animals to move the body dorsally or ventrally producing a sinusoidal locomotion pattern that preserves the side on which the animal is crawling.

In the laboratory, *C. elegans* is commonly cultured on the surface of agar plates seeded with a lawn of *E. coli* bacterium which serves as a food source. Due to its small size (approx 1 mm in length) many animals can be be grown on a standard size Petri dish. When dropped in liquid the animals can swim by rapid body bending with larger wavelength. It is sometimes described as thrashing behaviour.

C. elegans is one of the best characterized model organisms in biology. It was the first multicellular organism that had it's full genome sequence completed [3] and has a thriving community website *Wormbase* which provides a highly anno-

tated and detailed resource on its molecular biology [4]. *C. elegans* remains the only organism whose wiring diagram of all 302 neurons was mapped at electron microscopic resolution [5]. In addition, the developmental lineage of its 959 adult cells is known and allows researchers to look up a complete list of divisions that took place and lead to a specific cell of interest. An online resource *WormAtlas.org* stores this information and is a reliable resource for *C. elegans* anatomy. The versatility of *C. elegans* model organism allows studying severe defects of virtually paralysed worms. Mutants defective in fundamental neuronal molecules like synaptic vesicle proteins, voltage-gated channels and others can be cultured, referenced with rich resources of scientific data and analysed in the laboratory.

1.2 *C. elegans* genetics

The versatility of *C. elegans* for genetic experiments was shown perhaps most notably by Sidney Brenner in 1974. He screened populations of worms mutagenised with ethyl methanesulfonate (EMS) for morphological and locomotory defects [6]. In this study *C. elegans* males were used to perform complementation tests and mapping. Males occur rarely (about one in a thousand) but this rate can be increased by applying heat shock treatment. By mating males with the hermaphrodite a transfer of genetic markers from one hermaphrodite to another was achieved. Using this approach Brenner isolated and characterized 77 locomotion and morphology mutants: 56 uncoordinated, 14 dumpy and small, 1 long, 2 rolling and 4 blistered. These phenotypes were categorized based on manual observations and helped to identify concrete phenotypes linked with specific genetic alterations [6]. Since then many more phenotypes have been identified, their mutants cloned and characterized molecularly in great detail.

With the completion of the whole genome sequence of *C. elegans* [3] reverse genetics approaches started to be used. In reverse genetics a population of worms is mutagenised using psoralen and ultraviolet irradiation. Such approach is different to that of the EMS as it produces small to medium deletions. The population is subsequently screened with PCR primers designed specifically for the gene of interest. It takes about two months for a strain carrying a desired deletion to be constructed [7]. The *C. elegans* Gene Knockout Consortium has standardized this approach and provides it as a service to the research community.

The discovery of RNA interference using *C. elegans* added another powerful approach in studying genetic and molecular mechanisms of behaviour. Genes can be knocked down by feeding the nematodes with bacteria engineered to express double stranded RNA that is complementary to the gene of interest [8]. Systematic uptake of double-stranded RNA that triggers RNA interference is enabled by SID-1 transporter which can be found in most of *C. elegans* cells [9]. The cells that don't express SID-1 include neurons which limits this approach for studying molecular aspects of behaviour. A volume of research has gone into enabling the RNAi technology to work in neurons [10], [11], [12]. Ability to do so would not only allow much more rapid experiments and guide to a new generation of genetic screening - it would also create a big challenge to efficiently quantify produced phenotypes and interpret acquired data.

1.3 Behaviour analysis

Despite its simplicity *C. elegans* shows a wide array of behaviours such as feeding, mating, egg-laying, defecation, swimming, many subtle forms of locomotion on a solid surface and even nictating [13], [14], [2]. These forms of behaviour reflect the activity of the nervous system and are modulated by a range of factors such as external stimuli, neuronal structure and past experiences. A wide array of behavioural assays have been described and are continually created to allow probing diverse aspects of *C. elegans* biology.

Two major characteristics - morphology and locomotion are relied upon when studying spontaneous *C. elegans* behaviour. Changes in the cellular level of *C. elegans* can result in profound difference in regular or induced activity. Linking an observable phenotype with a genetic alteration allows elucidating complex neuronal processes. There are a few examples where observations in behaviour have helped unravel the molecular biology and explain the function of unknown genes. One example is in worms lacking an ion channel called *trpa-1*. The animals move well and seem healthy when inspected under the microscope [15]. However, the head of the animal performs an abnormally large head swing and is reproducible when observing a number of mutants. Quantifying the rate of head swings that cause a defect in foraging allowed guiding research that illuminated novel aspects of *trpa-1* function [15]. Similar observations have lead researchers

to identify novel phenotypes and get closer to unravelling the function of their corresponding genetic alterations.

1.4 High throughput analysis

In order to utilize the possibilities enabled by novel molecular tools it is necessary to develop methods that permit high throughput animal behaviour data collection and analysis. Behaviour is a complex process and to have sufficient statistical representation sufficient data coverage must be reached to ensure reliable data mining. Making sense of the data generated by these techniques is a challenging problem in itself. Most recent statistical methods and sophisticated machine learning algorithms have to be used to derive useful knowledge from large volumes of such data.

High throughput behavioural analysis of wide array of wild type and mutant *C. elegans* strains has the power of defining elusive behaviours and describing behaviours that would be very hard to tackle by eye. An example of such behaviours could be changes that are very gradual or can be observed only over a period of time (e.g. changes in the wavelength of locomotion or slight increases in locomotion velocity before events such as egg laying, etc.). Building such a platform would not only allow access to unprecedented amount of behavioural data but would also define a standard in the field which would enable comparing experiments done at different points in time, by different experimenters and in different laboratories. Accepted definitions of behavioural features of interest would allow collecting enough observations to reveal very subtle relationships between phenotypes that might reflect underlying genetic connections.

Many of these problems have been tackled over the last ten years in a form of automated animal tracking. However, the full potential of this approach has not yet been realised despite the increasing processing speed, memory capacity, sensors and even the resolution of digital cameras.

1.4.1 Animal tracking

C. elegans tracking is a process where the animal is followed and recorded over a period of time. It involves a digital camera, a light source and associated software to save the recoding of the experiment. Depending on the tracking resolution

motorized stages and various optics can be also part of the experimental setup. Such additions are needed in order to keep the animal in the field of view for the duration of the experiment. At the very top level tracker systems can be divided into two groups based on the resolution of tracking. Single worm trackers record animals at high spatial and temporal resolution and multi worm trackers tend to record multiple animals at once using lower magnification. The development of worm tracking closely follows the decreasing costs and increasing capability of computer technology and has already produced a set of very encouraging results that hint to the potential of such systems in the future.

The first automated worm tracker was developed by Dusenbery and used to study chemotaxis [16]. Despite using technology that is primitive by today's standards the platform was capable of tracking the movement of about 25 animals in real time at 1 Hz. *C. elegans* response to carbon dioxide and oxygen was surveyed as well as response to a variety of soluble forms of metals [17], [18]. The platform was programmed in BASIC09 and the later adoption of NIH Image software together with other improvements allowed tracking hundreds of animals [19].

Despite the promising results of the Dusenbery tracker it took several years before similar platforms were developed in other laboratories. A tracking system capable of tracking 50 animals was developed at Cori Bargmann's laboratory for the purpose of characterizing the role of neuropeptide Y receptor *NPR-1* in aggregation behaviour of *C. elegans* [20]. The platform used *DIAS* software that was developed for studying basic crawling behaviour of amoebiod cells and was adaptable for *C. elegans* data analysis [21].

Around the same time a tracking system that used computer-controlled motorized stage and a video camera mounted on a compound microscope was developed in Shawn Lockery's laboratory [22]. It focused on studying the role of pirouettes in chemotaxis behaviour and recorded position, speed and turning rate of individual animals. It recorded animals in high magnification at the rate of 1 Hz. The system used computer-controlled motorized stage to re-centre the animal as soon as it reached the edge of the field of view. With this approach experiments of moderate length could be recorded and analysed.

Another tracking system relying on computer-controlled motorized stage was developed at William Schafer's laboratory. This system allowed recording animal behaviour in high resolution and allowed researchers to refer to recorded data and

manually analyse egg-laying behaviour [23], [24]. It was further refined in partnership with Paul Sternberg's laboratory for automated collection and analysis of C. elegans locomotion data. The system was estimated to cost about 10,000 USD without the required compound microscope and related lighting and optics. It used custom algorithms developed in C Programming language that relied on OpenCV computer vision libraries. The recordings were done in high magnification and operated at 2 Hz [25], [26]. Even though the system was described as ready to use and intended for wider adoption in *C. elegans* community it was relatively expensive and hard to set up. Despite interest, it was not adopted by other laboratories and largely stayed within Schafer and Sternberg groups.

Several successful multi worm trackers have also been developed in a number of *C. elegans* laboratories. By sacrificing high resolution that can help identify subtle changes in *C. elegans* posture and locomotion multi worm trackers enable collecting larger amounts of data (more animals are recorded simultaneously) and allow studying it's social behaviours (such as aggregation, etc.). A system called *Nemo* was developed in Nektarios Tavernarakis's laboratory for two-dimensional object tracking for automated detection and analysis of *C. elegans locomotion*. By processing simple video files the tool allowed extracting, analysing and measuring nematode locomotion features [27].

Another platform called *The Parallel Worm Tracker* was developed in Miriam B. Goodman's laboratory. It used a simple thresholding algorithm to identify animals and worked both with bright field and dark field videos [28]. The system proved popular in the community due to the graphical user interface (GUI) that was also developed for the platform. Both *Nemo* and *The Parallel Worm Tracker* were developed in Matlab and allowed users to freely download the code and make custom alterations and additions to suit diverse needs.

The most sophisticated and detailed multi-worm platform to date has been developed in Rex Keer's group [29]. It was built as a high-throughput system that allows real-time analysis of behavioural data. The *Multi-Worm Tracker* can simultaneously quantify the behaviour of dozens of animals on a Petri dish at video rates. It requires minimal human effort and proved to be useful in studying locomotion, chemotaxis and habituation behaviours.

In addition to traditional solid surface tracking advances have been made in tracking *C. elegans* in liquid environments [30], [31], [32], [33]. Focusing on specific body regions and tracking body parts in conjunction with optogenetic

tools has also received a lot of interest and will be a valuable source of data in the future [34], [35], [36]. An excellent review of the history and current state of *C. elegans* tracking systems was done by Steven J. Husson *et. al.* and can be referred to for more detail [37].

1.4.2 Quantitative description of behavioural phenotypes

Early tracking platforms were simply recording animal behaviour and relied on manual analysis and annotation at a later date by the experimenter. However, to fully benefit from such systems automated algorithms based on computer vision and machine learning had to be developed. Such algorithms became available with increasing computational power and growing size of algorithm libraries that could be readily used depending on the programming language and development platforms.

Most of the automated worm tracking systems rely on similar data processing strategies when working with *C. elegans* behavioural data. They are commonly referred to as segmentation algorithms. The primary data is mostly in a form of a video recording. Therefore, the first step is to extract the individual pictures from each frame of the video file and identify the animal or multiple animals in them. This operation allocates all the pixels in the image to two groups - worm or background. The assignment is based on the intensity of each pixel - if it exceeds a certain threshold it is assigned to either worm or background. This operation is called thresholding and generates a two-colour binary image (black and white) that then is used to approximate the shape of the animal or animals. The next step usually is to abstract the shape even further and define a skeleton or spine of the animals. Such morphological skeleton of the animal shape runs from tail to head and is a most simple way to describe the shape and position of the animal at a given time point. Extraction of accurate skeleton is based on shape thinning algorithms [38] or most recently by identifying the points of higher curvature in the shape of the worm [39], [34].

The skeleton is represented as a collection of coordinate points and can be further subdivided into different biologically relevant segments. The collection of the coordinates allow a computation of various parameters such as the centre of mass (often referred to as the centroid), angles between two adjacent segments (measure of body curvature) and many others. The velocity of animals is com-

puted as the rate of change in the location of the centroid over time.

There are efforts to make a more universal system for animal identification. Multi-environment model estimation was used by Sznitman *et. al* to perform worm segmentation [39]. In this approach users have to manually identify the worm in the first frame of the video. A Gaussian mixture model is trained based on this selection to classify background and foreground pixels to identify animal shapes in subsequent frames. This system was able to extract swimming and crawling worms on agar as well as animals in microfluidic devices [39]. Other approaches based on Kalman filter [40] or recursive Bayesian filter [41] coupled with a worm model and utilizing inter-frame correlation have been used. Further development of these methods could lead to reliable segmentation of overlapping individuals and would be most useful in studying mating or social behaviours.

There are other considerations that determine the algorithms used to identify and abstract the animals. They range from the camera resolution, magnification used, the quality of the imaging conditions to number of animals on the plate. In case of multiple animals only moving objects are selected for further analysis. In multi worm trackers a special case is when animals collide and it becomes very difficult to distinguish one from another due to low magnification. In these cases new tracks are assigned to animals after they come into contact and then crawl to different directions.

Obtaining time series data of *C. elegans* outlines is just a first step in describing the phenotype of interest. Further abstraction and statistical analysis is needed to make biologically relevant interpretations. Fundamental behavioural features such as velocity (computed as a a change of identified shape position), reversals (negative velocity and sharp changes in animals path) and even shape (fitting ellipse to the outline of the animal) are computed using the outline data.

A promising complementary approach to extracting behavioural features has been described by Stephens *et. al.* [42]. In this approach each video is segmented and skeletonised using standard algorithms but instead of using the coordinates for the morphological skeleton the angles formed by the adjacent points along the skeleton are used. The angles are rotated by the overall mean angle and result in a position and orientation independent representation of the body posture over time. The covariance matrix of the rotated angles has a smooth structure and four eigenvectors can capture the majority (95%) of shape variance of animals crawling off food. This approach has been used by other groups and shown to

be effective in capturing the posture of even highly uncoordinated mutant worms [43].

1.5 Worm Tracker 2.0

William Schafer laboratory was one of the early adopters of the tracking technology to study molecular and genetic aspects of *C. elegans* behaviour. Despite the best efforts to build a platform that would be widely adopted by the *C. elegans* community the first attempt was largely unsuccessful. Nonetheless, valuable lessons were learned and with emergence of new technologies a new attempt was started to build a platform that would solve all of the previous issues. One requirement was that it had to be cost-effective that even laboratories not focusing on high throughput data analysis would be able to afford it. It also had to be very easy to assemble and use. In addition, associated software had to offer graphical user interface which would allow scientists without computer science training easily use the platform and derive satisfactory results. The platform also needed a proof of concept study which would illustrate the power of the system as well as help further the evidence of such platform's efficiency in unravelling complex biological processes represented by changes in animal behaviour.

In addition to Eviatar Yemini's efforts to build the tracking software and hardware for Worm Tracker 2.0 this thesis discuses a high throughput pipeline for efficient data analysis in Chapter 2. Behavioural data collected from more than 10,000 individuals is analysed and organized in a novel online *C. elegans* behavioural database which is presented in Chapter 3. The platform opens up a rich repository of behavioural data to the *C. elegans* community and proves the effectiveness of automated tacking and analysis. The data is freely downloadable for further inspection and processing. Chapter 4 describes how new behavioural features can be extracted from the data and novel *C. elegans* phenotypes identified from a collection of 350 strains. Chapter 4 focuses on navigation behaviour and studies reversals, turns and their intrinsic properties in greater detail.

The presented platform has a potential to become a widely adopted tool in *C. elegans* community. With standardized segmentation and feature extraction it could create a required repository for pooling *C. elegans* behavioural data obtained by different research groups. It would allow more reliable analysis as

well as considerably deepen the scope of data mining. It is an ongoing effort and with additional improvements described in Chapter 5 Worm Tracker 2.0 can become one of the fundamental tools in modern *C. elegans* neurobiology research.

CHAPTER 2

HIGH THROUGHPUT PIPELINE FOR *C. elegans* BEHAVIOUR ANALYSIS

Analysing a large number of experiments in an automated way requires reliable and efficient algorithms and tools. We have built a pipeline based on Matlab and PHP scripts with associated MySQL database to automate the extraction of behavioural data from more than 10,000 experiment recordings. In addition to the pipeline we have built an array of tools to guide the tracking process, view the results and share them in two online repositories that were linked with online *C. elegans* behaviour database.

Over a course of three years (2009 - 2012) we have recorded 16,117 *C. elegans* behaviour experiments and analysed them which resulted in creating 10.13 terabytes (TB) of data. Within this dataset 8.76 TB were raw experiment recordings with their corresponding annotation files and the results of the first pass of the analysis - video segmentation data and the second pass - normalized data. Normalized data organized *C. elegans* outlines for each frame in the video in a standardised format in preparation for behavioural feature extraction. The third pass of the analysis produced feature files which contained behavioural features extracted from the animal recordings. It amounted to 0.19 TB of data. In addition to the feature data we have also produced video files with segmentation overlay to enable easy illustration of segmentation algorithm accuracy. Overlay video files collectively add up to a 1.17 TB of data.

In order to efficiently maintain this large amount of data and keep track of analysed experiments we had to build a reliable pipeline for analysis (See Figure 2.1 for a schematic representation of the pipeline). Briefly, we have used 8 computers connected to their own worm trackers for recording *C. elegans* experiments (Figure 2.1 A). Recorded videos were saved locally on these computers and backed up on an external storage devices (NAS drives) (Figure 2.1 B). Each of the computers had Worm Analysis Toolbox installed - a software package that we have developed in the laboratory. Due to a highly parallel nature of the analysis we were able to run up to five toolbox instances to maximize the rate of analysis (Figure 2.1 C). Our analysis computers would be used for data collection during the day and analysis during the night. The toolbox workers connected to an internal relational MySQL database and queried table *segmentationExperimentList* to check which experiments remained to be analysed on the computer they were running (Figure 2.1 D). Raw segmentation data was saved locally as well as a second pass - normalized data - a precursor for feature extraction. Our software package outputs feature files as well as segmentation overlay videos. Both of these outputs were copied to NAS drives for security as well as to preserve space in tracker computers (Figure 2.1 E). Feature files were used for computing strain statistics and performing clustering analysis (Figure 2.1 G). Online *C. elegans* behaviour database used the results of aggregate strain statistics, feature and overlay files as well as annotations from the internal database to display the data for the scientific community on the Internet (Figure 2.1 H).

The pipeline could be easily controlled by re-running the instances of Analysis Toolboxes and resetting the analysis state in the *segmentationExperimentList* table. The methods bellow discuss various steps of the pipeline in more detail.

2.1 Data storage and backup

Backing up experimental data was scheduled frequently to ensure safe storage. We have used an open source backup program called DeltaCopy developed by Synametrics Technologies [44]. DeltaCopy is a Windows based tool which uses UNIX Rsync protocol to allow incremental backup (it copies only the part of the file that is modified rather than copying all of it once again). It saved the time needed for backup as well as bandwidth and served as an excellent tool for large

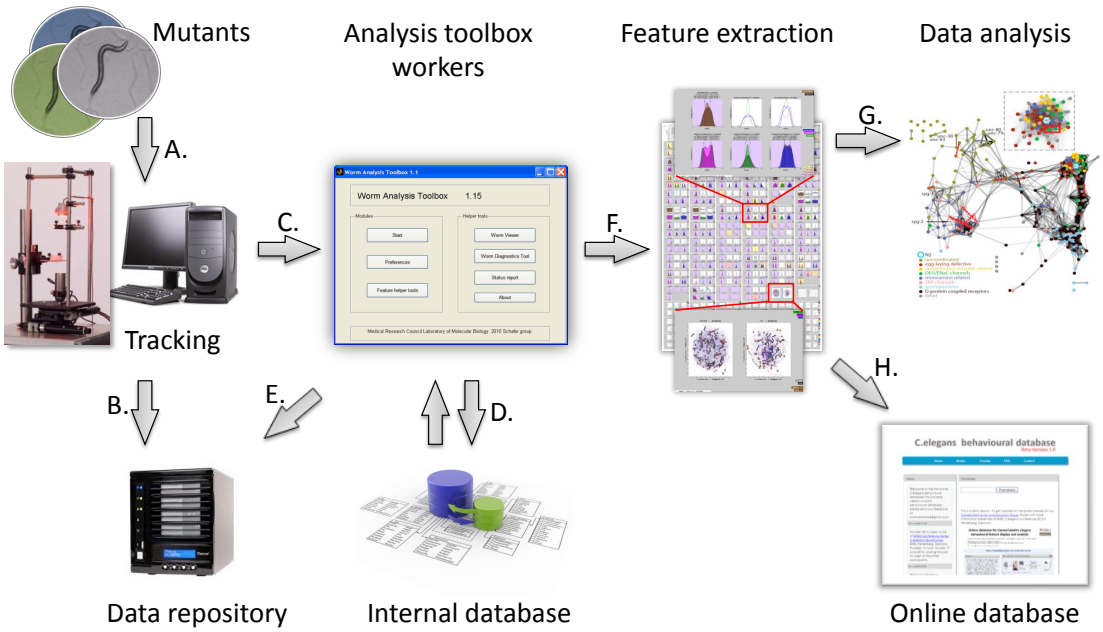


Figure 2.1: Pipeline. Schematic representation of high-throughput *C. elegans* behavioural analysis pipeline.

backup jobs. In addition, DeltaCopy supports task scheduler, email notifications and easy restore in case of erroneous changes.

Alternative storage such as cloud has not been explored, however, in setups that would contain significantly larger number of tracking and analysis units such approach would be worth exploring. Especially because analysis could be run on the cloud as well.

A collection of 329 strains and all of the analysed data amounted to approximately 10.13 terabytes (TB) of data. A very quick calculation indicates that for 18,000 mutants 554.3 TB of storage space would be need. In this case cloud storage and big data processing would be necessary. An in-house solution taking advantage of Hadoop and Hadoop Distributed File System (HDFS) would be needed or a similar solution on the cloud like Amazon EC2.

2.2 MySQL database for experiment data

To index experiment recordings and manage analysis we have built an internal relational database framework using MySQL. We used Matlab scripts to connect to the database and perform table access, insert and update operations. We used Java-based data access technology (JDBC driver) to enable Matlab and MySQL connection. Please refer to Figure 2.2 for database schema.

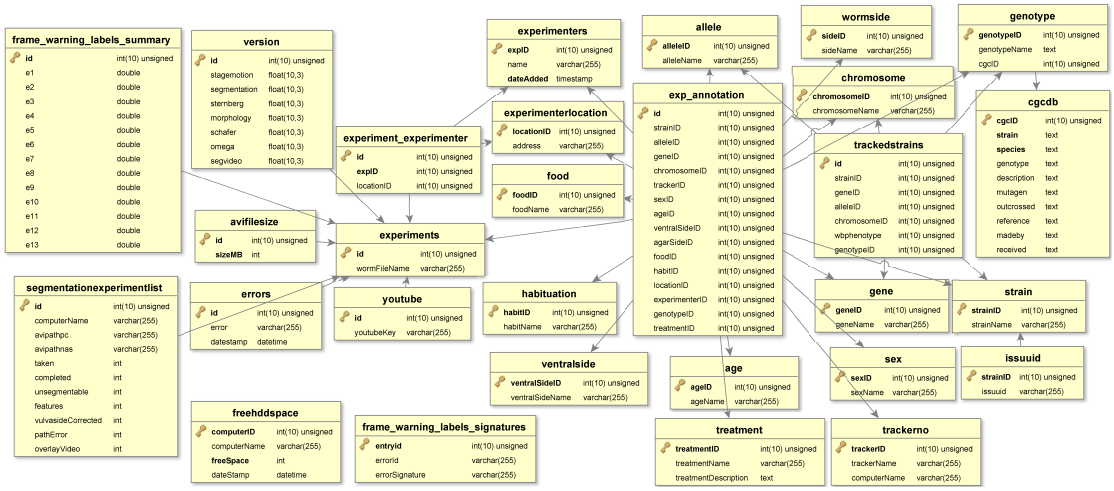


Figure 2.2: Database schema. We have used a relational database to index experiments, control the analysis progress and store their annotation data. Please refer to Appendix A for more details on the database tables.

There are three essential database tables that form the central logic of the pipeline: *experiments*, *segmentationExperimentList* and *exp_annotation*. Briefly, table *experiments* contains an index of all experiments tracked in the laboratory. Each experiment receives a unique identification number which is used to facilitate relations to other database tables. Table *segmentationExperimentList* serves as a central table for experiment analysis process. Worm Analysis Toolbox software connects to this table and searches for all unfinished experiments on the client computer and proceeds with the analysis. Table *exp_annotation* collects all of the annotation information for each experiment. Tables such as *gene*, *allele*, *chromosome*, *habituation* and others contain detailed annotations of each tracked experiment and are referenced in *exp_annotation*. This table is also used

when serving the data to the online database. For full list of tables and table descriptions refer to Appendix A, Table A.1.

2.3 Worm Analysis Toolbox

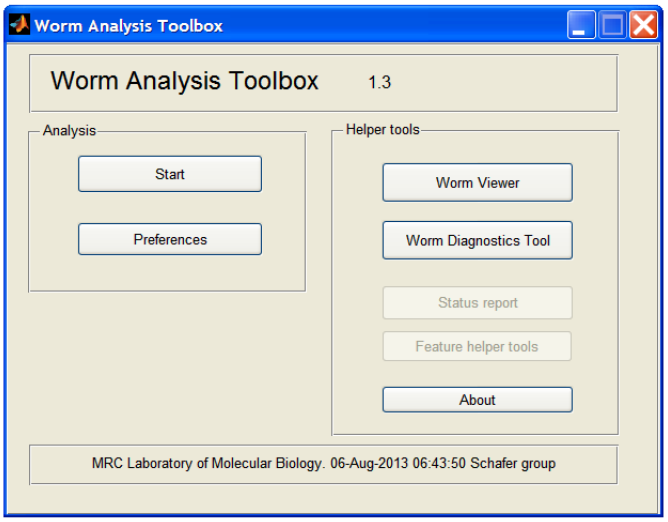


Figure 2.3: Worm Analysis Toolbox. Main window of Worm Analysis Toolbox graphical user interface.

To analyse *C. elegans* behaviour recordings we have built a custom software using Matlab language and environment for technical computing. We have called the software package Worm Analysis Toolbox as it contains a graphical user interface and allows biologists without computer science training to effortlessly use it (Figure 2.3). The toolbox is compiled using Matlab Compiler. It allows running it without the need of a Matlab software license which some of the laboratories might not have or might not be willing to invest in. Instead, in addition to Worm Analysis Toolbox we distribute Matlab Compiler Runtime (MCR) which when installed allows running the toolbox without the need of Matlab software license. It enables users to use the full benefits of our software freely. Installation instructions can be found online at *http://www.mrc-lmb.cam.ac.uk/wormtracker/index.php?action=analysis*.

The toolbox operates in two modes - database mode or standalone mode. In the database mode it connects to the analysis pipeline database and enables high throughput analysis. It connects to the MySQL table *segmentationExperi-*

mentList and looks for unfinished experiments in the computer that it is running on. Due to the fact that our tracker computers had multiple core processors and the nature of the analysis is very parallel we were able to run up to five toolbox instances to speed up the analysis. To avoid toolbox workers analysing the same experiment simultaneously a MySQL table *segmentationExperimentList* had a column *taken* which was updated when the experiment was selected to be analysed by the toolbox. This way of scheduling allowed making sure that all of the toolbox instances are analysing different experiments up to the point where no unfinished experiments are left on the computer. In the standalone mode, the analysis toolbox takes a directory of interest, searches for complete experiments and saves the results locally after analysing them in a sequential way.

Preferences section of the toolbox allows switching the toolbox between these two modes. In addition, it allows customizing the analysis process. Other tools included in the toolbox allow checking the performance of the segmentation algorithms and enable users to view the time series of the feature data for each individual experiment. Worm Analysis Toolbox provides a complete package that researchers can use to facilitate detailed behavioural analysis using the methods that we have developed in out laboratory. More details on the other sections of the toolbox are described below.

2.3.1 Preferences

Preference section of the analysis toolbox is divided into three parts. Top left part contains the main analysis steps, bottom left allows calibrating automatic head and tail detection and right part enables control of preferences related to analysis pipeline (Figure 2.4). Analysis status for each experiment is saved and when the toolbox is executed again it continues from where it left. Options *Redo stage motion diff*, *Redo stage motion detection*, *Redo segmentation*, *Redo head and tail detection* and *Redo worm normalization* allow the user to re-compute various steps of the analysis. This functionality is especially useful when newer versions of the toolbox are released on our website. *Output video* option instructs the toolbox to save a segmentation overlay video to allow manual checking of animal identification accuracy. *Remove overlay video* deletes the video after it is produced to save space. *Calculate features* outputs behavioural features in a Matlab file. Feature files for all analysed experiments were saved in a *.mat* files

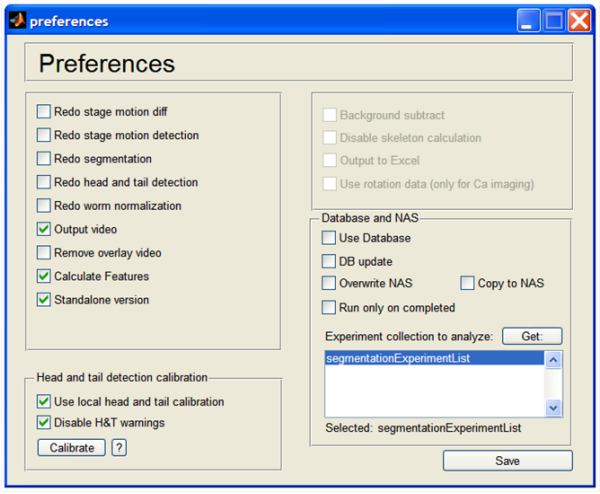


Figure 2.4: Preferences window. Preferences section of Worm Analysis Toolbox can be used to customize analysis steps.

with a setting allowing them to be Octave compatible. Octave is a high-level interpreted language primarily intended for numerical computations and is very similar to Matlab. It is a free software under GNU General Public License (GPL) and is published by the Free Software Foundation. It allows other scientists to use the feature data freely without the need of Matlab license. The *Standalone version* property is a switch to use the toolbox independently of our pipeline.

The pipeline is controlled by the *Database and NAS* panel on the left (Figure 2.4). It controls database access mode, backup to NAS drive and analysis for only already analysed experiments. Most importantly it allows selecting a database table which contains experiment collection of interest. Tables with the same structure but different experiments can be created to separate into different analysis runs or represent different behavioural experiment groups. This section in the preferences allows instructing the pipeline to work on the experiment list of interest.

Worm Analysis Toolbox uses machine learning to predict head and tail of the animal (for more details see Chapter 3.3.7). The algorithm analyses motion and colouration data together with a training set assembled in our laboratory to identify the head and tail of the animal. In cases where behaviour experiment recordings have different magnification and lighting settings to those used in our laboratory the efficiency of the algorithm might decrease. Head and tail detection calibration allows to build a local training set used for automated head and tail

detection. More information about the head and tail detection calibration step can be found in Appendix A.

2.3.2 Video diagnostics tool

Declaration note: Segmentation algorithm visualisation was developed by Eviatar Yemini and is used in the video diagnostics tool.

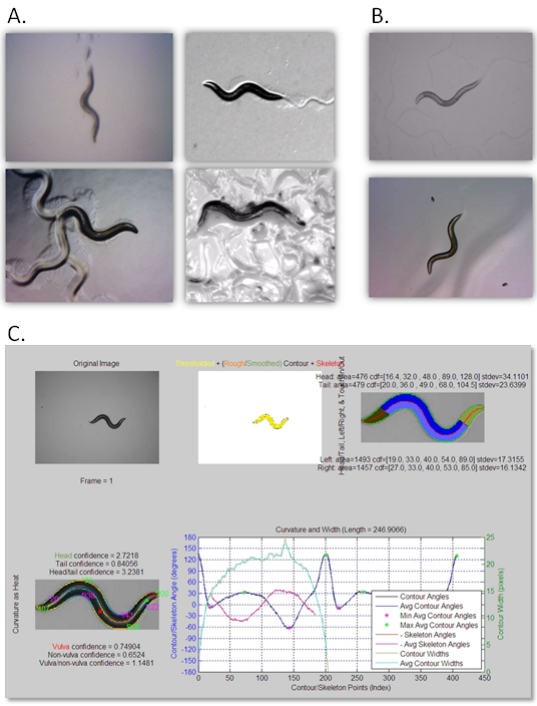


Figure 2.5: Video diagnostics tool. **A**. Example of frames unsuitable for reliable segmentation. **B**. Example of frames suitable for reliable segmentation. **C**. Video diagnostics toolbox view for a selected frame that helps experimenters fine tune recording conditions to ensure reliable segmentation algorithm performance.

A tool has been built to guide experimenters in achieving best possible video recordings to ensure high quality feature data (Figure 2.5). The best way to obtain reliable feature data is to ensure the quality of *C. elegans* behaviour recordings. Diagnostics tool illustrates the inner workings of the segmentation algorithms and allow fine-tuning the data collection.

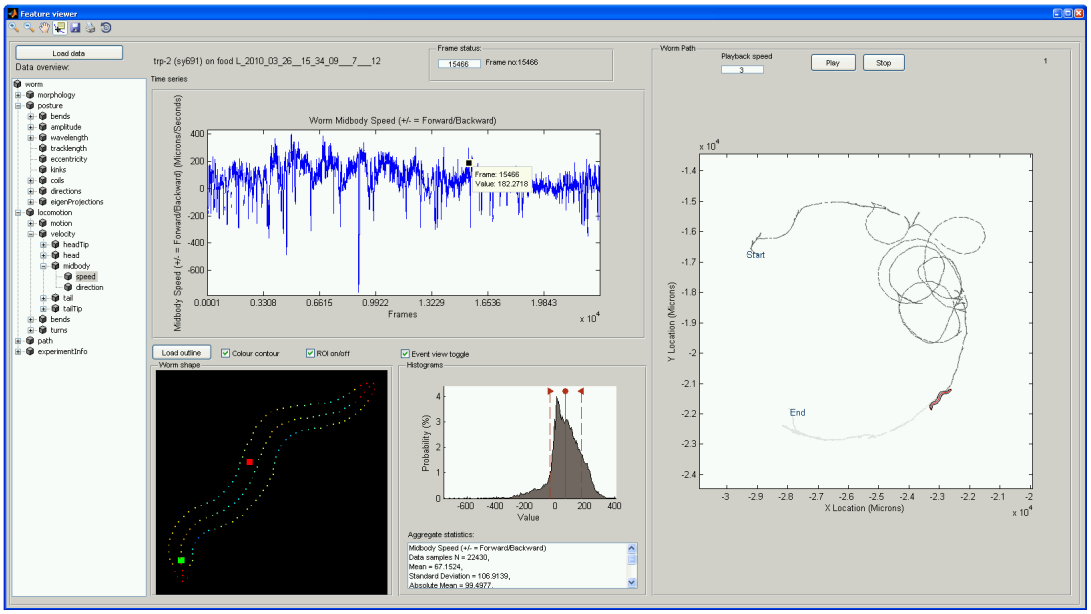


Figure 2.6: Worm viewer. Five panels allow users to explore extracted behavioural data for experiment of interest. Users can select specific feature or explore the annotation of the experiment. Time series data of the selected feature is plotted for further interpretation. The outline of the animal at a selected point in the experiment is displayed together with the associated histogram. The full path of the animal is reconstructed and the location of the animal at a chosen point is indicated.

2.3.3 Worm viewer

Worm Viewer is a tool for researchers to load behavioural experiment features extracted by Worm Analysis Toolbox. It allows selecting any of the computed feature measures and viewing their time series data. The viewer is split in to five main panels that display the feature list, time series traces, worm path, outline and histogram of selected feature. Each of these panels will be briefly described in more detail.

Data overview. Users can load a feature file of interest using the *Load data* button. Available data will be displayed in a selection box on the right hand side of the viewer window. The selection box is implemented using Java libraries from within Matlab and allows expanding the data structure of the feature file. More detailed description of the features can be found in Chapter 3.

Time series. This panel plots the time series of the feature of interest selected in the data overview box. The plot has the X and Y axes annotated to assist the interpretation of the plots. The user can use zoom, hand or data cursor tools at

the upper left corner of the viewer window to inspect the time series plot further. While using the data cursor to select a point on the time series plot other panels will be refreshed and will display the data associated with that point.

Worm shape. This panel displays the shape of the worm at a selected point in the experiment. In case the feature file is loaded without the normalized files only the morphological skeleton of the animal shape will be displayed. The full shape can be loaded and displayed by pressing the *Load outline* button. It will look for necessary files relative to the loaded feature file. In case they can't be found it will prompt the user to specify the location. ROI toggle allows users to see the magnified view of the shape of the animal instead of the whole field of view as it was recorded during the experiment.

Histograms. The panel shows the histogram of selected feature time series and will print the aggregate statistics that can be copied to another program manually for further statistical analysis.

Worm path. Full path of the worm over the course of the experiment is displayed. A shape of the animal at a selected point is also displayed in this panel. Darker path shows the path already travelled by the worm and the lighter path is the path it will travel in the remaining part of the experiment. The user can playback the experiment using the buttons at the upper part of this panel. Other panels will refresh while the playback is on. Playback speed can be specified by the user.

2.4 Online tools

Online *C. elegans* behaviour database (*www.wormbehavior.mrc-lmb.cam.ac.uk*) was developed using PHP (a server-side scripting language) with a MySQL database back-end. PHP scripts generate the web pages after retrieving the data from the database. The location and identification number for files that were associated with specific strains presented in the online database were stored in MySQL tables and retrieved by the website scripts. Several essential application programming interfaces (API's) were used to interface with external services that allowed efficient data presentation. We have used Issuu and YouTube platforms to serve aggregate statistics PDF files and segmentation overlay videos respectively. Both of the platforms have excellent viewers that can be embedded in our

online database pages. Such quality and functionality would be well beyond what can be achieved without large engineering teams. In addition we have served approximately 30GB worth of PDF data and 1.17 TB of video data. It would be a big technical challenging to enable reliable and rapid access to such an amount of data independently. Therefore, using both of these online platforms allowed efficient delivery and presentation of our results. Both platforms and the way we have used them are described in more detail below.

Issuu is an online service that allows realistic and customizable viewing of digitally uploaded material, such as portfolios, books, magazine issues, newspapers, and other print media. It has a versatile and convenient reader that can be embedded in other website. By using Issuu API we have uploaded the PDF files of aggregate statistics of 305 *C. elegans* mutants and embedded the viewer in our online database pages. Each of the files received unique Issuu identification numbers which was saved in our local database. By using this identification key Issuu viewer could be used and linked to specific PDF files. Since the files are large (100 MB) and the viewer loads each page as it is being accessed the users have very short waiting times and can interactively explore aggregate behavioural data of each of the 305 presented strains.

YouTube is a video-sharing website on which users can upload, view and share videos. It allows users to upload videos freely and has a highly functional and robust viewer that can be embedded in external websites. We have developed a script that takes a directory of overlay videos, uploads them to YouTube using it's API and saves the unique YouTube ID to our database for embedding in online behaviour database pages. In addition of having efficient way of serving 1.17 TB of video data to the scientific community we can use rich YouTube channel usage analytics to survey the impact of our research.

FTP server at MRC Laboratory of Molecular Biology was used to store feature data that is also presented on the strain pages of online behaviour database. Users can log in using anonymous account and copy the data freely. It allows for other research groups to take the whole data and run their own analysis just like a novel analysis described in Chapter 4.

Wormtracker Website was built to share the Worm analysis Toolbox and the tracker software. In addition a detailed hardware setup instructions as well as related publications are presented there as well. The website was programmed using PHP language and can be found on *http://www.mrc-lmb.cam.ac.uk/wormtracker/*.

2.5 Acknowledgements

I would like to thank Eviatar Yemini and Andrè Brown for useful discussion on the pipeline and the tools. I would like to thank Tobias Klöpper for advice on database development and management.

CHAPTER 3

A DATABASE OF *C. elegans* BEHAVIOURAL PHENOTYPES

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

Declaration. *The work presented here came out in a paper while writing this thesis [45] and was carried out in collaboration with other team members of the Schafer group. Laura Grundy collected experimental recordings that were used for the analysis. Andrè Brown helped to supervise the work, discussed and guided the development of algorithms and helped compile the figures and text which featured in the paper. Eviatar Yemini has built the tracker software which enabled saving the video recordings of the animals, designed the tracker hardware and participated in developing core algorithms for the analysis. He also carried out several steps of statistical data analysis and helped to write the paper.*

A principal output of the nervous system is motor behaviour. Therefore, a variety of neural perturbations ultimately manifest as changes in motion, making locomotion a useful phenotype for neurogenetics in model organisms. For the nematode *C. elegans*, the study of mutations that cause visible defects in spontaneous crawling has given insight into diverse neural functions. Despite these successes, extending behavioural phenotyping to large-scale screens remains a challenge. In fact, 85% of *C. elegans* genes have no reported phenotypic effect when knocked down using RNAi, though most knockdowns do detectably reduce fitness over several generations [46]. These findings suggest that there is a phenotyping gap: knocking down most genes has an effect, but one too subtle to see by manual observation.

For this phenotyping gap to be closed, it is desirable to have a system that is capable of measuring phenotypes both extensively and intensively [47]. For behaviour, extensive phenotyping requires measuring many parameters that quantify motion, posture and path as well as the frequencies and intervals between relevant behaviours such as reversals and sharp turns. Achieving intensive sampling requires following individual worms with high temporal resolution over extended intervals. These conditions can be met with single-worm trackers that follow freely behaving worms and that use a motorized stage to keep the worm in the camera's field of view [37].

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

formatic approaches that link complex phenotypes with genetic perturbations. For neurogeneticists, the summary statistics and visualizations make it possible to identify behavioural phenotypes in mutants of interest.

3.1 Worm Tracker 2.0 and the Online Database

We collected data with Worm Tracker 2.0 (WT2), which uses a mobile camera to automatically track and record single worms (Figure 4.11 A). Worms were tracked on food on an immobile platform isolated from tracking motion. Videos were analysed to extract the worm's contour and skeleton (Figure 3.1 B). The head was identified automatically from the image, and the dorsal-ventral orientation was obtained from user input. We further reduced contour and skeleton data by extracting relevant features [48] using automated methods developed for use in previously described worm trackers [49], [26], [38], [42], [50]. We developed several new algorithms for feature extraction to quantify parameters related to motion state, crawling, foraging, dwelling and exploratory range (Appendix B). At the top level, the extracted features include measures of morphology, posture, locomotion and path dynamics (Figure 3.5 for illustrations and Appendix B, Table B.2 for the full list of measurements). These primary features were also evaluated in different contexts to give more complex parameterisation: for example, mean speed was measured over the entire video as well as independently for periods when the animal was moving either forward or backward. Likewise, dorsal and ventral bending were measured over the entire body and in specific regions such as the head, tail and midbody. Finally, specific behavioural events such as reversals or omega turns were used to generate secondary parameters, such as the frequency, time spent in execution and distance covered during the event. All together, the permutation of these measures yielded 702 distinct feature measurements. By running eight tracking units in tandem, we could achieve a throughput of approximately 128 recordings (8 recordings each of 15 strains plus 18 wild-type controls) per day. The data from each individual are available as a time series of features. The WT2 feature viewer provides a visual assessment of these feature time series, frame by frame, accompanied by the worm's contour, skeleton and location on plate (Figure 3.1 E). Each video is also provided, via YouTube, with an overlay of the skeleton, contour, head, ventral side

and annotations for unsegmented frames (Figure 3.1 B). These two distinct views provide complementary ways to verify the fidelity of the features. Furthermore, aggregate measurements and statistics per strain are available directly on the website, presented graphically in PDF files, in HDF5-compatible MAT files, and as spreadsheet-compatible CSV files. Inclusion of the skeleton data in the MAT files permits new features to be rapidly computed from the existing data set.

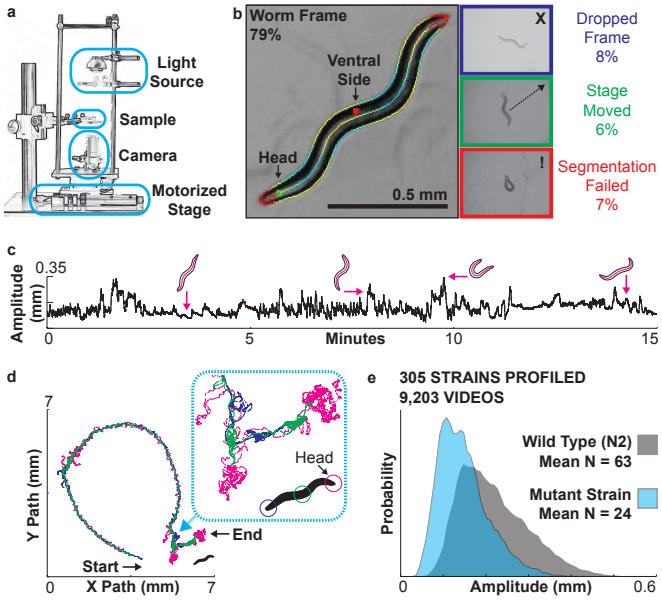


Figure 3.1: The Worm Tracker 2.0 (WT2) system and phenomic database. **A**. Schematic of the tracking hardware. **B**. Videos are segmented to extract the worm contour and skeleton for phenotypic analysis. Frames from which contour and skeleton cannot be extracted are annotated with a colour-coded border. The border indicates whether the frame was dropped (because of the processor's inability to keep up with incoming video frames), a stage movement occurred (blurring the worm image) or worm extraction failed (owing to an unrecognized shape); features are not computed from such frames. Numbers indicate the percentage of frames in each category. **C**. Single-worm behavioural time series measuring maximum amplitude, defined relative to the major axis of the best-fit ellipse. Several shapes are shown with their corresponding amplitudes. **D**. Single-worm path (same worm as in **C**). **E**. Feature histogram measuring maximum amplitude (for the strain presented in **C** and **D**).

3.2 Behavioural phenotyping resuls

We compared each strain in the database to the lab-stock wild-type N2 strain. Because of month-to-month variability in the N2 data (Appendix B, Figure B.2),

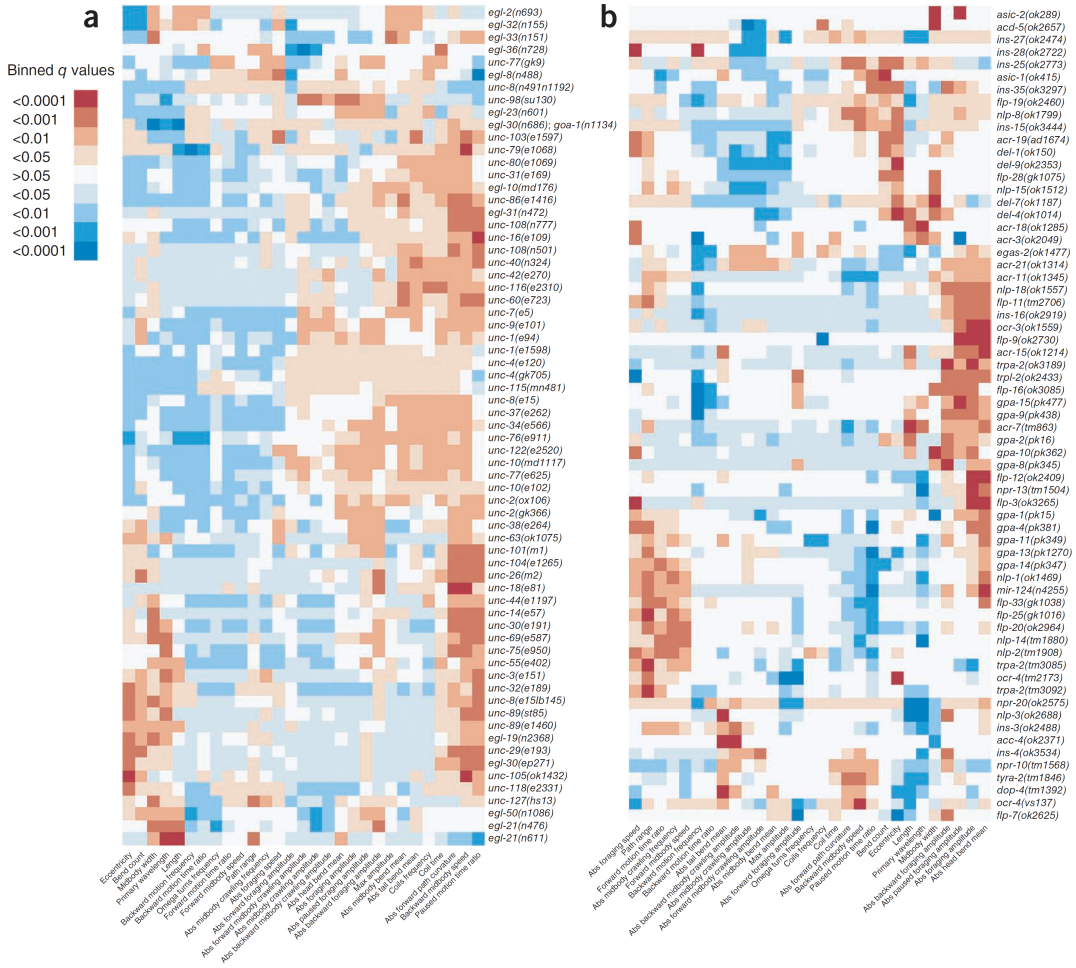


Figure 3.2: Phenotypic summaries using selected features for subsets of strains. Colours in the heat maps are used to indicate the q value for each feature for the comparison between each of the mutant strains listed on the right and the N2 reference data. Red values indicate features that have a significantly higher value in the mutant, whereas blue indicates significantly lower values in the mutant. Genes and features were both hierarchically clustered for easier comparison. **A**. Subset of mutant strains with previously known locomotion phenotypes. **B**. Subset of mutant strains with no previously reported locomotion phenotype. Abs, absolute; max, maximum.

each mutant strain was compared to an N2 control data set of videos collected within a 2-week window centred around the date the mutant data were collected (meaning within 7 d before or after the mutant recording). When a measure was detected exclusively in either the strain or its control (for example, some strains never perform reversals), we used Fisher's exact test to assess significance. For the remaining cases, we used the Wilcoxon rank-sum test, controlling the false discovery rate with a q value in place of P [51]. All strains were significantly

different from wild type at $q = 0.05$, including those representing 76 genes with no previously characterized phenotype.

The feature measurements in the database appear to provide a useful description for a wide range of behavioural phenotypes. For visibly uncoordinated mutants (Figure 3.2 A), the measured locomotion features serve as a sensitive fingerprint for phenotypic similarity. Loss-of-function alleles of the same gene (for example, of *unc-2*, *egl-21*, *unc-4*, *unc-108* and *unc-89*) exhibit highly similar phenotypic signatures, as do mutants affecting genes encoding components of known molecular complexes (for example, *unc-7/unc-9* ([52]), *unc-79/unc-80* ([53]) and *unc-38/unc-63* ([54])). It is reasonable to suppose that the signatures of genetically uncharacterised mutants (such as *egl-31* and *egl-33*) might provide insight into the molecular or cellular functions of the mutant gene.

With regard to phenotypes identified in previously uncharacterised knockout strains (Figure 3.2 B), the paucity of multiple knockout alleles makes attribution of phenotypes to many gene deletions provisional. Nonetheless, some genes are represented by multiple alleles, including the previously uncharacterised *trpa-2*, a member of the TRPA family of cation channels that function in nociception and thermosensation in many animals including *C. elegans* [15]. From our analysis of three *trpa-2* deletion mutants (*ok3189*, *tm3085* and *tm3092*), we determined that all three showed a similar significantly altered posture during reversals (Appendix B, Figure B.3). These results suggest a potential proprioceptive role for *trpa-2*, a hypothesis that can be investigated further. We also detected new locomotion phenotypes for multiple alleles of several genes previously implicated in other processes (Appendix B, Figure B.3), including the TRPV channel *OCR-4* ([55]) and the TRPC channel *TRP-2* ([56]). These results show that automated tracking can detect previously unnoticed locomotion phenotypes, even in relatively well-studied strains.

In addition to heatmaps we have visualised the distances between strains along selected features in a form of a box plot. Figures 3.3 and 3.4 display the distribution of strain differences to wild-type (N2) along selected feature dimensions. The median of Wilcoxon rank-sum similarity test for each selected feature is presented in a *-log10* scale along with 75% and 25% quantiles. The box plot whiskers are computed to indicate strains that fall outside of the 95% range. The entire list of q-values for each strain and for each feature is available online in a supplementary table of the paper forming the basis of this chapter [45]. Supplementary tables

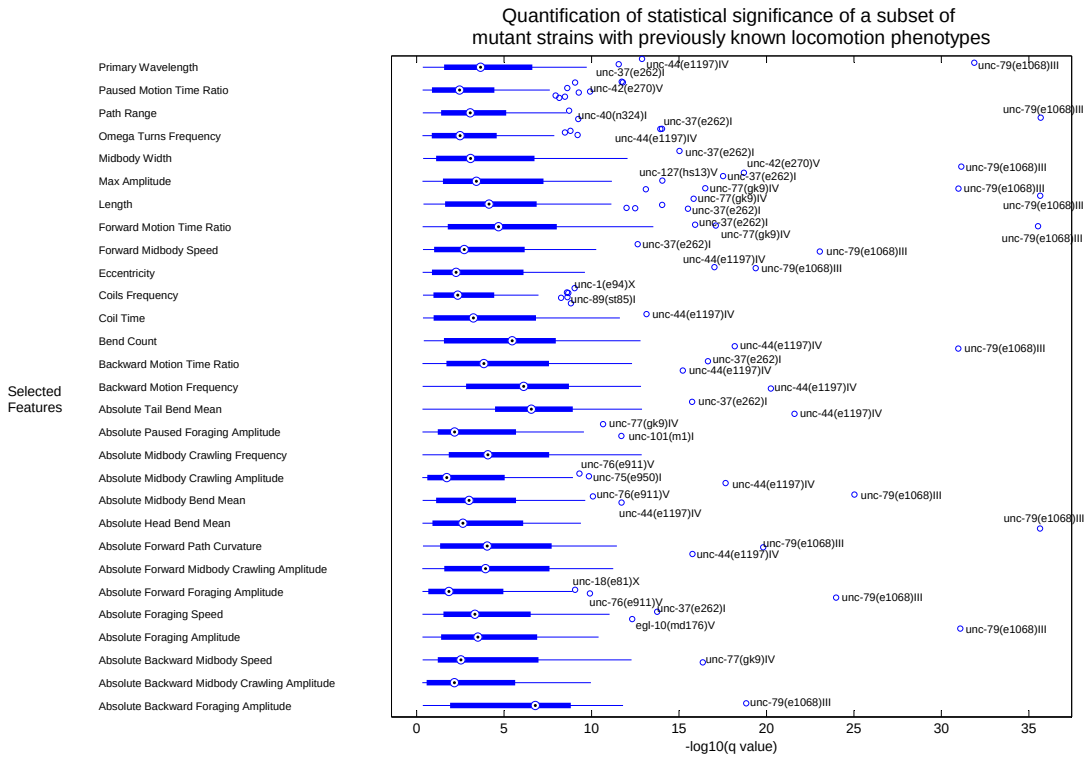


Figure 3.3: Phenotypic summaries displayed in a box plot for a subset of mutant strains with previously known locomotion phenotypes. Wilcoxon rank-sum test with control for false discovery rate was used to compute q-values. Boxplot indicates q-values that lie outside of the 95% confidence interval assuming a normal distribution which results in the whisker value of 0.9529. (*Note that the distributions for the selected features are not normal and this assumption is used only to display the results.*)

(especially tables 4 and 5) can be found online and freely downloaded for further analysis from:
http://www.nature.com/nmeth/journal/v10/n9/full/nmeth.2560.html#supplementary-information

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

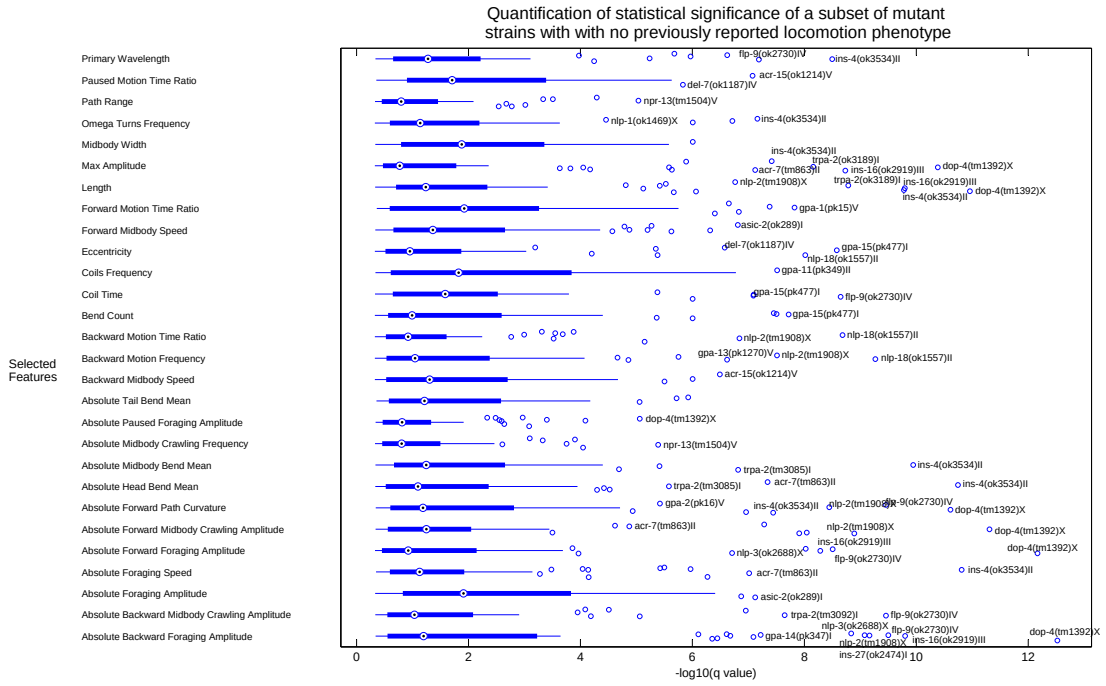


Figure 3.4: Phenotypic summaries displayed in a box plot for a subset of mutant strains with no previously reported locomotion phenotype. Wilcoxon rank-sum test with control for false discovery rate was used to compute q-values. Box-plot indicates q-values that lie outside of the 95% confidence interval assuming a normal distribution which results in the whisker value of 0.9529. (*Note that the distributions for the selected features are not normal and this assumption is used only to display the results.*)

their own approaches to studying the behaviour of particular mutant strains.

We intend to add more strains to the phenotype database and to increase the richness of the data available for each genotype. One approach would be to record each strain under a variety of environmental conditions, probing different sensory modalities and behavioural responses. A recently published study of behavioural responses to thermal stimuli highlights the potential effectiveness of this approach [57]. It should also be possible to extract additional features from the existing video data. This could involve developing new algorithms for feature measurement or more unsupervised approaches based on time-series motifs [43]. It is our hope that making these data available to the community will encourage other researchers to contribute collaboratively to the emerging science of behavioural informatics.

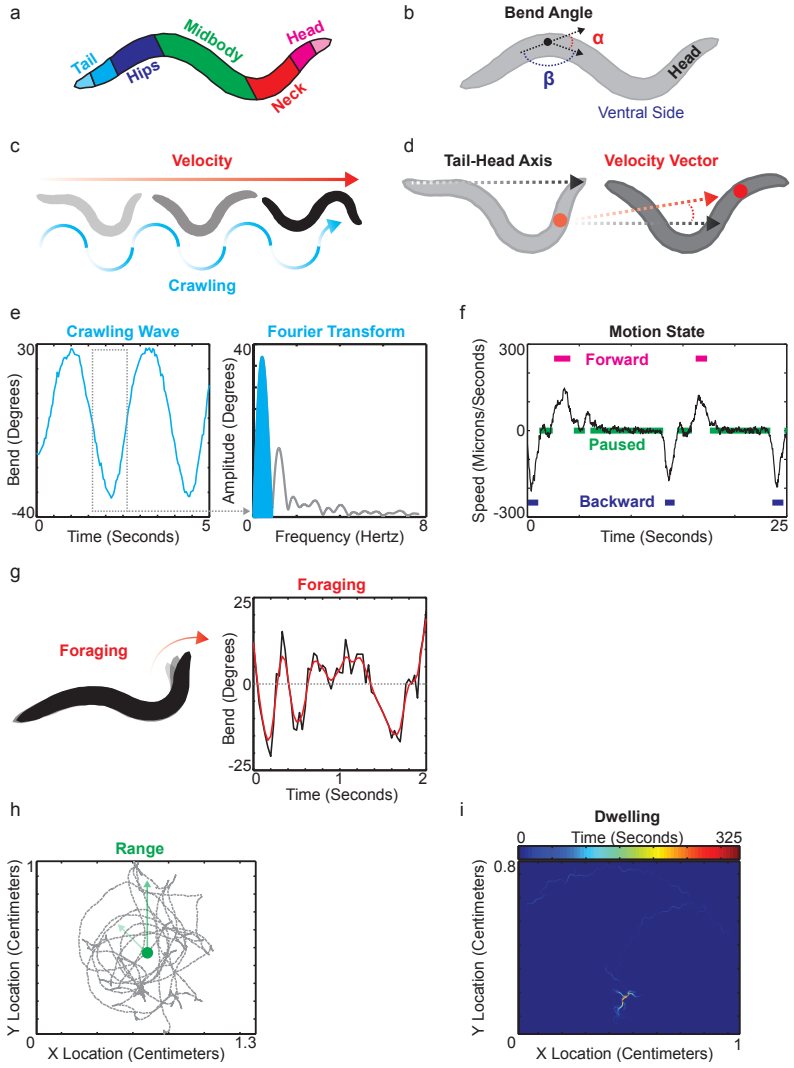


Figure 3.5: Feature computation.. (Next page).

3.3 Methods

3.3.1 Collection of worm video data

Worms were maintained under strictly controlled conditions up until the point of tracking [58]. To ensure sufficient phenotypic sampling, we filmed at least 20 young-adult hermaphrodites per strain for 15 min that were spontaneously behaving on food. The camera magnification was set to between 3.5 and 4.5 m per pixel (a corresponding field of view (FOV) of approximately 2.5 x 2 mm^2 at 640 x 480 resolution) and the frame rate was set to 20-30 frames per second to

Figure 3.5: (Previous page.) **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 (α) is the difference in tangent angles at each point; or, alternatively phrased, the supplementary angle (α) with respect to the angle formed by any three consecutive points (β). 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.

ensure a high-resolution analysis. We observed a 30-min wait, before tracking, to allow worms to habituate after being picked and moved to their tracking plate. A list of all recorded strains is provided in Appendix B, Table B.1.

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

3.3.2 Dorsal-ventral annotation and head-tail detection

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

3.3.3 Feature normalization

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

3.3.4 Computer code

Unless otherwise noted, the image processing, statistics and bioinformatics functions were available through common Java libraries (v.1.6), standard Matlab toolboxes (v.2010a) or standard R packages (v.2.15) or were created as needed. Several publicly available MathWorks packages significantly facilitated the analysis and figures within this paper: the *videoIO* toolbox by G. Dalley, *swtest* function by A.B. Saida, *fexact* function by M. Boedigheimer, *export_fig* function by O.

Woodford, *notBoxPlot* function by R. Campbell and *rdir* function by G. Brown.

3.3.5 Worm Tracker 2 details and algorithms

The Worm Tracker 2.0 (WT2) hardware guide and free software are available at: *http://www.mrc-lmb.cam.ac.uk/wormtracker/*.
In summary, a camera, illumination and motorized stage are combined to follow a single worm navigating a thin bacterial food lawn on an agar Petri dish. The software uses a closed loop wherein live video is used to guide the stage and keep the worm centred in the camera's FOV. The Petri dish housing the worm is immobile while the camera, stage and illumination move as one. Therefore, the worm is isolated from external forces such as stage movement.
The camera is a DinoLite AM413T with zoom magnification. We used a 640 x 480-resolution, 30-f.p.s. camera with a magnification that results in 3.5-4.5 μm per pixel and an FOV of roughly 2.5 x 2 mm^2 at a focal distance of nearly 1 cm. Older videos have a frame rate of 20-30 f.p.s. New videos maintain 30 f.p.s. The illumination is a red, Philips Lumileds, side-emitting Luxeon III Star (LXHL-FD3C). An opal diffuser provides roughly uniform lighting over the FOV. The wavelength is 627 nm to avoid exciting a short-wavelength avoidance response through the *C. elegans LITE-1* receptor. Finally, the motorized stage uses Zaber T-NA08A50 linear actuators. The stage travels 5 cm at up to 8 mm/s in orthogonal x and y axes with a resolution of just under 0.05 μm.
Worms were maintained as previously described in [58]. The protocol is also available online *www.mrc-lmb.cam.ac.uk/wormtracker/webcontent/trackingProtocol.pdf*. Briefly, worms are maintained at room temperature, approximately 22 °C. All plates are fresh, having been poured within 1 week of use and kept at 4 °C until roughly 24 h before use. Strains are maintained on standard NGM plates [6] seeded with three drops of OP50. Six adult worms are transferred to a new plate to maintain stocks. At least two generations are passaged in these conditions before tracking. The evening before tracking, at roughly 5 p.m., L4 hermaphrodites are picked to a fresh plate, with ten worms per plate. The next morning, 3.5-cm low-peptone NGM plates are seeded with 20 μl of OP50, in the centre of the plate, and allowed to dry. This 20μl drop of OP50 is nearly circular and roughly 8 mm in diameter. The L4 worms from the evening before, now young adults, are transferred to the centre of the food on the 3.5-cm plates, one worm to a

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

3.3.6 Worm segmentation

Video frames were extracted using the Matlab *videoIO toolbox* by G. Dalley. The toolbox is a Matlab wrapper for a DirectShow or ffmpeg (depending on the platform - Windows or Linux - of the user). It allows accessing video files in Matlab using fast and stable native multimedia framework.

There is a sharp contrast between the worm and background in our video images. Worm pixels were segmented from the background using the Otsu method [59] to find a threshold. The largest eight-connected component in the thresholded image was assumed to be the worm. Frames in which the worm touched the image boundaries, was too small, lacked a clear head and tail or had unrealistic body proportions were not analysed further. Frames containing stage movement were also removed to eliminate bad segmentations wherein the worm image may be blurred (see below: **Absolute coordinates**). Given our desire for accurate and precise measures as well as the large data volume (due to a high video frame rate), we erred on the side of caution and attempted to reject ambiguous segmentations rather than include them. Once the worm was thresholded, its contour was extracted by tracing the worm's perimeter. The head and tail were located as sharp, convex angles on either side of the contour. The skeleton was extracted by tracing the midline of the contour from head to tail. During this process, widths and angles were measured at each skeleton point to be used later for feature computation. At each skeleton point, the width was measured as the distance between opposing contour points that determine the skeleton midline. Similarly, each skeleton point served as a vertex to a bend and was assigned the supplementary angle to this bend (Figure 3.5 B). The supplementary angle can also be expressed as the difference in tangent angles at the skeleton point. This angle provides an intuitive measurement. Straight, unbent worms have an angle

of 0°. Right angles are 90°. And the largest angle theoretically possible, a worm bending back on itself, would measure 180°. The angle is signed to provide the bends' dorsal-ventral orientation. When the worm has its ventral side concave within the bend, the bending angle is signed negatively. Pixel count is a poor measure of skeleton and contour lengths. For this reason, we used chain-code lengths [60]. Each laterally connected pixel was counted as 1. Each diagonally connected pixel was counted as $\sqrt{2}$. The supplementary angle was determined per skeleton point using edges 1/12 the skeleton's chain-code length in opposing directions along the skeleton. As there are an insufficient number of skeleton points to determine the angles at the ends of the skeleton, these angles were undefined (i.e., the first and last 1/12th of the skeleton had no bending angle defined). One-twelfth of the skeleton (Figure 3.5 A illustrates division of the worm body into parts) has been shown to effectively measure worm bending in previous trackers and likely reflects constraints of the body-wall muscles and their innervation and cuticular rigidity [26].

3.3.7 Ventral side annotation and head detection

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

The head was detected in each chunk of oriented frames. The head and neck perform more lateral motion (for example, foraging) even in uncoordinated mutants. Therefore, we measured lateral motion at both worm end points, across each chunk-unless the chunk was shorter than one-sixth of 1 s, which is too short to reliably measure such motion. In our setup, the head allows more light through than the tail and therefore appears lighter (has higher mean intensity).

Therefore, we also measured the grayscale intensity at both worm end points, across each chunk. Linear discriminant analysis (LDA) was used on a combination of lateral motion and intensity at the worm end points for a training set of 68 randomly chosen videos. This classifier was then used for the entire data set to automatically detect and label the worm's head.

3.3.8 Absolute coordinates

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

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

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

3.3.9 Feature overview

All feature formulas were computed from the worm's segmented contour and skeleton. The skeleton and each side of the contour were scaled down to 49 points for feature computation. Wild-type worms have four quadrants of longitudinal, staggered body-wall muscles [61]. Each quadrant contains 24 such muscles with the exception of the ventral-left quadrant, which has 23. With a sampling of 49 points, the skeleton and contour sides have a well-defined midpoint. Moreover, because the worm is confined to two dimensions, its body-wall muscles present roughly 24 degrees of freedom (although in practice it seems to be far less [42]). With 49 points we had two samples per degree of freedom and, therefore, expected to sample above the Nyquist rate for worm posture.

A common notation was used to define the body parts (Figure 3.5 A). The head is controlled by the first four body-wall muscles per quadrant-approximately one-sixth the length of the worm [5]. Similarly, the neck is controlled by the next four body-wall muscles per quadrant-also approximately one-sixth the length of the worm. For this reason, we defined the head as the first one-sixth of the worm and the neck as the next one-sixth of the worm (skeleton points 1-8 and 9-16, respectively). For symmetry, we defined the tail and 'hips' in a similar manner, on the opposite end of the worm. The tail was the last one-sixth of the worm and the hips were defined as the next one-sixth (skeleton points 42-49 and 34-41, respectively). The midbody was defined as the remaining middle one-third of the worm (skeleton points 17-33). For some features, the head and tail were further subdivided to extract their tips: the first and last 1/12 of the worm (skeleton points 1-4 and 46-49), respectively.

Frame-by-frame features were represented by top-level histograms and statistics as well as subdivisions exploring their values during forward, backward and paused states. This was to measure behaviours that depend on the state of motion, such as foraging amplitude, which is reduced during reversals in wild-type worms [62]. Many features were signed to reflect dorsal-ventral orientation, forward-backward trajectory and other special cases (for example, eigenworm projection) to capture any asymmetry. Finally, event-style features (coiling, turning, and motion states) were summarized using global and local measures. Global measures include the event frequency, the ratio of time spent within the event to the total experiment time, and a similar measure for the ratio of the dis-

tance covered within the event to the total distance travelled by the worm (when available). Local measures include the time spent in every individual event, the distance covered in each event (when available), and both the time and distance covered between each pair of successive events. See the Appendix B for further details.

3.4 Acknowledgments

We thank C. Cronin and P. Sternberg for providing the code from their single-worm tracker, which serves as the basis for elements of our WT2 analysis GUI and phenotypic features cited from their publication; J. Lasenby and N. Kingsbury for guidance in computer vision; A. Deonarine for useful discussions; and R. Samworth, S. Chavali, G. Chalancon, S. Teichmann and M. Babu for guidance in statistics and bioinformatics techniques.

CHAPTER 4

Quantitative analysis of *C.elegans* navigation behaviour

The nematode *Caenorhabditis elegans* uses turns and reversals to respond to a variety of environmental changes and efficiently explore its surroundings. Reversals and turns produce pirouette behaviour that interrupts forward locomotion and reorients the animals upward in the gradient during chemotaxis and is observed in thermotaxis and aerotaxis. The behaviour of the nematode during reorientation is studied to shed light on the mechanisms of navigation. However, such analysis had limited scope due to the lack of high resolution data for wild type and mutant strains. Here we use a novel *C. elegans* behaviour dataset from online *C. elegans* behavioural database to precisely identify turning behaviour that is characterized by rapid change in the direction of locomotion. In addition to omega turns we have identified a subset of other turns that enable similar reorientation and are equally important in navigation. Quantitative analysis of turning velocity, turn duration and body posture of 327.7 hours of high spatial and temporal resolution wild type behaviour allowed better understanding of reorientation mechanisms and their dimensionality. Both omega and other turns were analysed in the context of correlation to reversals and turning ventral/dorsal bias drawing a reliable baseline of the wild type phenotype. It was compared to a database of 350 *C. elegans*

mutants representing a wide variety of genes. A collection of strains showing significantly different behaviour from wild type were identified and provide a unique insight into the molecular aspects of *C. elegans* navigation.

Ability to navigate in an unknown environment is one of the most fundamental behaviours in animals. Efficient navigation relies on integrating sensory information with previous experiences to produce adaptive behaviour. Neural circuits and underlying molecular mechanisms govern animals' locomotion behaviour in navigation. One of the most common models for dissecting molecular aspects of behaviour is a nematode *Caenorhabditis elegans*. It is one of the few species that has a completely defined synaptic wiring diagram of its nervous system that consists of 302 neurons [5]. Extensive work has been carried out to characterize the escape behaviour circuit of *C. elegans* [63]. In addition, neuronal control of locomotion during exploratory behaviour has also been rigorously studied [64]. Finally, animals response and navigation along the gradients in thermotaxis and chemotaxis have been described in detail [65], [66], [22].
The studies suggest that *C. elegans* navigation displays characteristics similar to those of a biased random walk. Periods of relatively straight movement **runs** are perturbed by **turns** facilitating rapid change in the direction of locomotion. Such rapid direction changes that involve reversals and turns are called **pirouettes** and have been shown to play a fundamental role in the taxis behaviour [22]. Most common turns in pirouettes are called **omega turns** because the posture of its body during the turn resembles a Greek letter omega. Omega bends are characterized by steep body bends where the animal touches or nearly touches a part of its own body [67], [68]. Mechanisms using pirouettes have been shown both experimentally and computationally to be a primary strategy for *C. elegans* chemotaxis in laboratory assays [69]. However, the dimensionality of turns, their shape, ventral or dorsal bias, correlation with reversals and their length have not been precisely defined in large wild type datasets or compared extensively across mutant libraries.
We sought to precisely quantify aspects of nematode navigation and identify strains that show different statistics to that of the wild type. We have built a turn detection algorithm based on intuitive measure of rapid direction of locomotion change and focused on four aspects of *C. elegans* reorientation behaviour:

turn correlation with reversals, turn ventral bias, turns followed by long reversals (resembling pronounced escape behaviour) and finally overall reversal rates for mutants. We have identified 6416 turns in 327.7 hours of wild type behavioural data. A significant fraction (1873) of these turns did not contain omega bends but showed similar reorientation capabilities and were included in phenotypic analysis. Quantitative analysis of turning velocity, turn duration and body posture allowed establishing a reliable description of turning behaviour. Phenotypic analysis of 350 strains showed a wide array of mutants significantly different from the wild type strain in terms of reversal correlation, ventral bias and overall reversal rates. This data draws up a promising list of candidates for laboratory experiments which will aid the efforts to define molecular aspects of *C. elegans* navigation. In addition this analysis illustrates how *C. elegans* behavioural database can be used to shed more light on important aspects of the nematode biology and can allow unprecedented exploratory analysis to guide experimental work in the lab.

4.1 Materials and Methods

4.1.1 Data and strains

Behavioural data for this work has been used from the *C. elegans* behavioural database. It is available online and full set of features can be downloaded from designated database repository directory. Matlab based algorithms were developed to load individual feature files and extract necessary data. Each strain contains a collection experiments that imaged different individuals from the strain population for 15 minutes in high spatial and temporal resolution. The frame rate for the tracking experiments was approximately 30 frames per second enabling detailed analysis of the behaviour. Majority of experiments were 15 minute long except some which reached up to one hour of continuous recording. Freely moving worms were imaged using a standard tracking protocol described in the database. The strains were obtained from CGC consortium and cultured in standard laboratory conditions. A collection of 350 strains were tracked and analyzed. The coverage of strains differed and ranged from 8 minutes of cumulative data for off food experiments for ocr-3 (ok1559, RB1374) to 46.6895 hours for unc-7 (e5, CB5). Wild type experiments of N2 strain were most extensively covered comprising a dataset of 327.7 hours.

4.1.2 Turn detection

Omega bend detection was one of the original feature extraction steps while creating *C. elegans* behavioural database. The algorithm was based on omega bend detection algorithm described in [50]. It is important to point out that the exact algorithm as it is described in [50] could not be replicated due to the fact that it took advantage of coiled shape segmentation. Coiled shape segmentation was not included in this iteration of *C. elegans* behavioural database and will be part of future analysis.

Therefore, the algorithm for omega bend detection used for the current database was developed to loosely follow the underlying principle of the algorithm described in [50]. The algorithm attempted to identify a bend travelling across the animal's body as it executes a turn. The head of the animal would bend, it would approach the centre of its body and the wave would travel through the body to the tail resulting in high curvature in the tail at the end of the turn. This approach was replicated in the algorithm used for the omega bend extraction for the database and was presented in the paper discussed in Chapter 3, [45]. However, without a comprehensive segmentation method for coiled shapes the algorithm based on curvature of the animal failed to reach desired accuracy.

New turn detection algorithm was implemented based on the direction of locomotion. Direction of locomotion was obtained by computing the angle between x-axis and a vector formed by point on the head and the tail of the animal. For illustration please refer to Figure 4.1.

The first step of the new algorithm was to compute a direction of locomotion angle derivative. An example can be seen in Figure 4.2A. The peaks of the derivative correspond to rapid changes in direction and can be used to identify turning behaviour. The peaks were identified by selecting the rate of turning that is larger than 1.5 degrees per second. This threshold was chosen based on observation, trial and error and empirical data. The beginning and the end of turns were estimated by taking the points at a zero crossing of direction angle derivative. Identified turns that were 1/10 of a second apart were joined and treated as one long turn because the beginning and the end of such adjacent turns would be virtually indistinguishable.

In order to illustrate that the new algorithm is able to identify turning events better a random sample of one hundred wild-type (N2) *C. elegans* experiments

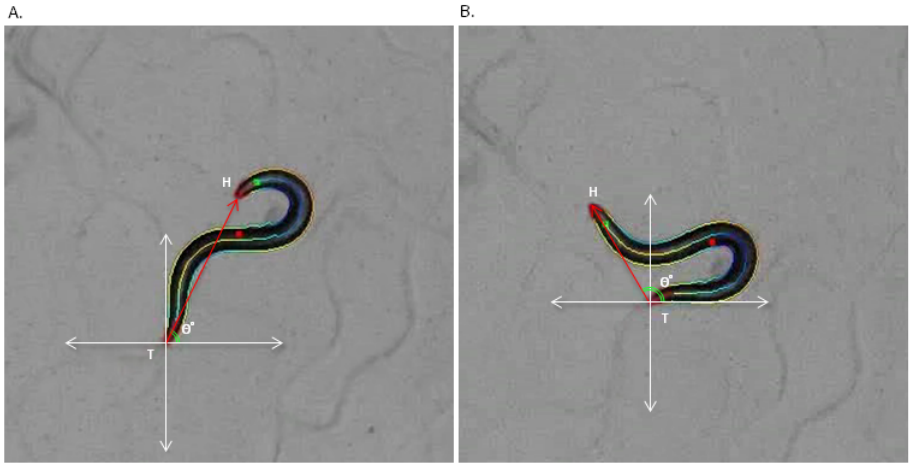


Figure 4.1: Tail to head direction vector angle. Head and tail direction vector is indicated in red, H stands for head and T for tail. The red dot in the middle of the animal indicates vulva side which was manually annotated by the user when recording the video. White arrows are the axes along which the angle θ will be calculated. **A**. The worm is initiating a turn with a deep head bend. The angle θ starts changing rapidly at the beginning of the turn. **B**. The worm slides the head along the ventral side of the body and exits by moving forward resulting in a sharp locomotion direction change. The angle θ has rapidly changed significantly over a short period of time.

was selected. Both of the algorithms were executed and results compared. Figure 4.3 illustrates the accuracy of the old and the new algorithms. New algorithm was able to correctly identify 376 turning events and outperformed old algorithm which was able to correctly identify only 171 turns. It also managed to maintain a false positive rate of only 5.2897%. Turning events were wrongly identified only in cases where the animal was deeply covered in food or touched another object thus momentarily interfering with head and tail detection algorithm. Further optimisations of the new algorithm that would deal with head and tail errors would yield even lower false positive rates. In addition to that, further study of false negatives would increase the accuracy of the algorithm even further. Both of these steps are planned for the next iteration of the algorithm.

4.1.3 Turn characteristics

Omega turn detection. Omega turns were identified by finding the turns where the worm touched its body or its head came close to its tail indicating a deep bend. Distance from head to tail was calculated for the duration of the turn. If

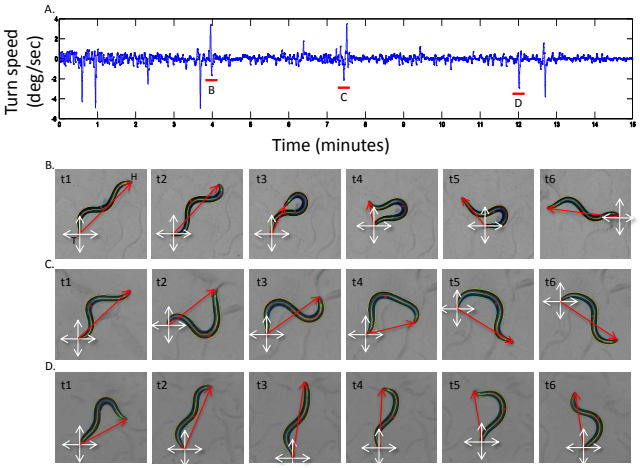


Figure 4.2: Reorientation rate and three different turn examples. The difference in change of head to tail vector angle allows identification of rapid angle change events. These events correspond to the worm reorienting itself. Three events were chosen to illustrate most common ways the animal reorients itself. **A**. Reorientation rate plot for a 15 minute long experiment. The spikes indicate rapid changes in direction of the locomotion. Three examples are indicated to illustrate possible turn behaviour. Snapshots from *t1* to *t6* show the worm assuming different shapes as it is making a turn. **B**. *Standard omega bend*, worm bends the head at *t2*, brings it close to the midbody occasionally touching it at *t3*, turns the head outward to exit the turn making a Greek letter Ω like shape at *t4* and exits the turn at *t5*, *t6*. **C**. *Wide turn*. Worm reverses at *t1* and *t2*. Turns the head first dorsally (*t3*) then ventrally (*t4*). Exits the turn widely without bringing the head close to the tail at *t5*, *t6*. **D**. *Pivoting turn*. Worm reverses at *t1* and *t2*, turns head dorsally at *t3* and continues forward locomotion at *t4*, *t5* and *t6*. Due to the rapid succession of these steps the change of direction of locomotion allows identifying these turns.

the distance between head and tail reached less than 20% of the animals body length the turn was classified as having a deep omega like bend.

Other turns. The remaining turns had a rapid change in direction of locomotion but did not have the head come close to the tail over the duration of the turn.

Turn speed. Turn speed was calculated by taking the maximum rate of change in direction of locomotion over the duration of the turn.

Velocity before the turn. The maximum velocity of the midbody of the animal within one second before the turn was calculated for each of the events. It was extracted for both forward and backward locomotion.

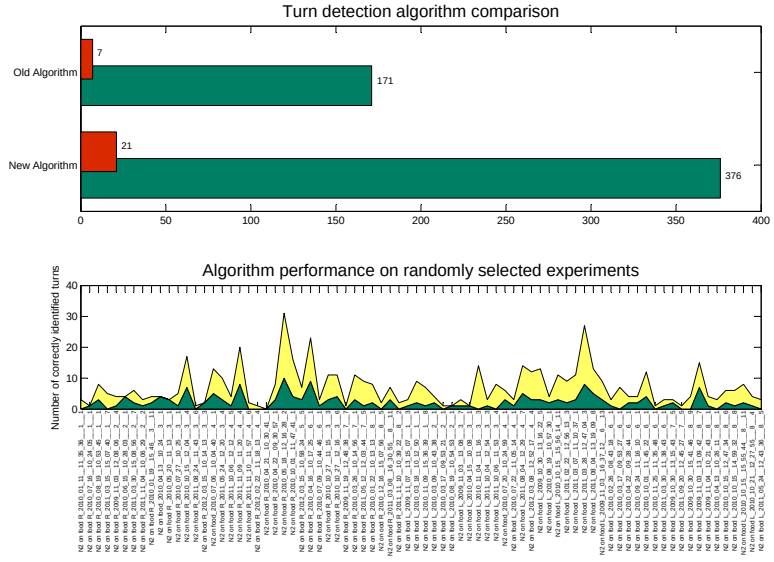


Figure 4.3: Turn detection algorithm comparison. Two turn detection algorithms were compared for ability to find turning events and accuracy. One hundred random N2 videos were selected and both algorithms were used to identify turns. Old algorithm was implemented in a similar fashion to [50] and used in [45] which forms the basis of the *C. elegans* behavioural database. A new algorithm looking for changes in direction of locomotion and is much more sensitive in detecting turning events (**top panel**). False positives occurred in both algorithms mainly due to the animal being obscured by food or other large objects. **Bottom panel** shows the performance of both algorithms for each of the files used in the test. In each of the cases the new algorithm outperformed the old one. Both of the algorithms collectively identified 78 experiments which contained turns.

Body curvature. Curvature of the animals head, middle and tail was computed by taking a mean of 1/6th of the front of the animal for head, 1/6th of the end of the animal for tail and 2/3rds of the animal in the middle. The maximum curvature of each three of these body parts during one second before the turn was extracted.

Reversals before turns. Similar to [50] a window of 5 seconds was inspected before the beginning of a turn to determine if there was a reversal preceding the turn. The distance in microns for each reversal was saved to distinguish between short reversals and long reversals.

Long reversals were defined as reversals longer than the body length of the animal (approx. 1000 microns).

Reversal rate. Mean of the reversal rate was computed by counting the number of reversal events for each minute of the experiments.

4.1.4 Statistical analysis

Pearson's Chi-Square test of independence was used for categorical data. For velocity and body posture features Students' t-test was used [70]. For both of the significance tests Bonferroni corrected p-values were used to account for independent tests that were made comparing 330 strains to wild type. Error bars for categorical data were computed using 95 percent bootstrap confidence interval [71], [72].

4.2 Results

4.2.1 Wild type analysis

The wild type dataset analysis consisted of Schafer lab young adult N2 strains tracked over a period of three years. It consisted of 1288 recordings of which the majority were 15 minute long. There were 25 experiments of 2 hours of continuous recording with it the total wild type data added up to 327.7 hours of spontaneous worm locomotion. Each of the experiments were recorded at high magnification and high temporal resolution. The recording frame rate ranged from 30 frames per second to at least 25 frames per second. Different frame rate depended on the system load and lighting conditions. The dataset analysis resulted in detection of 6416 turn events that were extensively checked for their accuracy.

4.2.2 Different turn types

Out of 6416 turns 4543 turns satisfied the criteria of turns containing omega bends [64]. These turns contained events where animal touched its own body or had a near miss. The remaining 1873 turns had turning rate as high as one found in omega turns but showed animals assuming different body postures Figure 4.4 C, D. One of the most central functions of the turn is to reorient the worm to allow alternate direction of locomotion. Turn magnitude in degrees for omega

turns and other turns was computed in Figure 4.4. The figure shows the angle by which the animal changed it's direction of locomotion. Negative angle means it made a ventral turns, positive means dorsal. Wild type preference for ventral turns is evident. The mean of turn magnitude for omega ventral turns is 152° with standard deviation (SD) of 52°. Omega dorsal turn magnitude mean was 143° with SD of 58°. Other turns show similar characteristics - ventral mean is 116° and SD of 46° and dorsal turn magnitude mean is 107° and SD 47°. This data shows that the animals are able to reorient efficiently while making turns that don't contain omega bends. It brings light to the fact that turning can take place in various other shapes and forms and that these other turns are also extremely important for *C. elegans* navigation. Includig these other turns into the analysis of nematode navigation can lead to better description and understanding of the underlying molecular and neural processes.

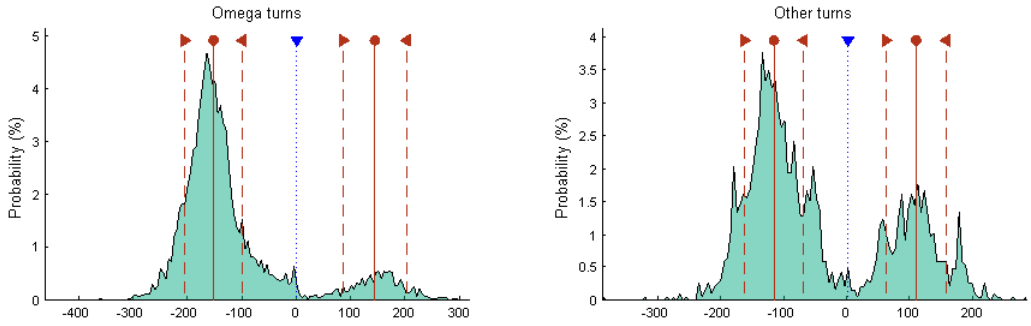


Figure 4.4: Turn reorientation magnitude. Turn reorientation magnitude for wild type strain extracted from 327.7 hours of behavioural data. Total number of turns detected: 6416. Difference in angle of locomotion before and after the turn presented in degrees. **Omega turns**. A total of 4543 omega turns were identified. Ventral reorientation mean is 152° and standard deviation is 52°. Dorsal turns were less numerous and had a mean of 143° and standard deviation 58°. **Other turns** A total of 1873 other turns were identified. Ventral reorientation mean is 116° and standard deviation is 46°. Dorsal turns were less numerous and had a mean of 107° and standard deviation 47°.

4.2.3 Turn characteristics

Extensive wild type dataset allowed detailed turn characteristic analysis, Figure 4.8. Duration of turns with omega bend show a mean of 4.5±2 (± SD) seconds. Other turn duration is similar to that of omega bends. Turns that took more than 10 seconds were events where the animals turned continuously several or

more times until they continued forward locomotion.
No single preferred rate of turning could be identified in omega turns Figure 4.8 B.1. The range of values from 1.5 deg/sec to 5 deg/sec could be achieved with equal probability. It indicates that turning behaviour is a complex process and the animals vary the rate at which they reorient. Other turns show less variable turn rates that mostly do not exceed 4 deg/sec.
Maximum midbody velocity for the time period of one second before the turn was extracted illustrating the locomotion activity before the turn. Two clear modalities were identified showing that the animal was either moving backward or forward before the turn. Backward velocity mean of 345±112 microns/second (μ/sec) was identified which was larger than that of forward velocity - 201±81 μ/sec. When making other turns backward velocity is also larger than forward - 311±138 μ/sec for backward and 172±102 μ/sec for forward. *C. elegans* velocity before the turn resembles normal distribution for forward and backward locomotion and shows characteristics of rapid reversal that is faster than the forward locomotion. These results have to be interpreted with caution because the precise start of the turn is hard to define. The start of turns in this data was chosen to be the point when the rate of turning peaked, however, for different turning scenarios that can be either before the reversal, during or after (or not correlated with reversal at all). Therefore, the two modalities here show that worms tend to be moving at these two velocities before turns with specified probabilities.
Body curvature before the turn showed a maximum head bend angle to be larger when the worm turns ventrally (58±14 degrees for ventral turns and 39±13 degrees for dorsal). When making omega bend the head bending was larger for both ventral and dorsal turns (other turn ventral 39° ± 13° , dorsal mean 36° ± 11°). The distributions of maximum ventral and dorsal head bend angles in other turns were more similar than the ones in omega turns. Midbody and tail maximum bending before the turn was very similar differing only in the probability of dorsal ventral sides depending on the orientation of the worm before the turn. These probability distributions also have to be interpreted in context of highly fluid start of the turn conditions.

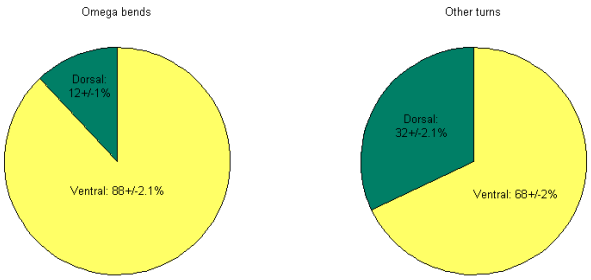


Figure 4.5: Wild type (N2) turn ventral bias. Turn ventral bias analysis based on 4543 omega turns and 1873 other turns of wild type strain. The majority of turns show ventral bias. When making omega turns ventral bias was (88±2.1%) and when making other turns (68±2%). Error bars were calculated using bootstrapping with 95% confidence interval.

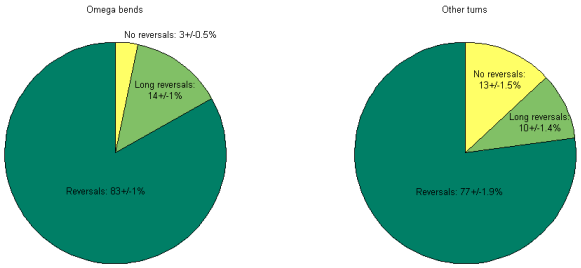


Figure 4.6: Wild type (N2) turn correlation with reversals. Turns are correlated with reversals. Only 3+/-0.5% of omega like turns show no correlation with reversals. A fraction of 13±1.5% other turns indicate no correlation with reversals. Long turns were also identified and show similar fraction - 14±1.5% for omega and 10±1.4% for other turns.

4.2.4 Turn ventral bias, turn reversal correlation and reversal rates

We have used the wild type dataset to inspect whether the turns involved contraction of the dorsal or ventral body muscles. We have observed a strong ventral bias for omega turns 88±2.1% and a smaller ventral bias for other turns 68±2% (Figure 4.5). It confirms previously observed bias for wild type animals to turn ventrally nearly all the time when making omega turns [50]. We have also compared turning and reversal correlation in wild type animals, Figure 4.6. The data shows 97±1% of omega bend containing turns to be positively correlated with reversals. Our analysis also confirms high correlation rates previously shown in literature [64], [50]. Other turns show slightly smaller percentage of turns corre-

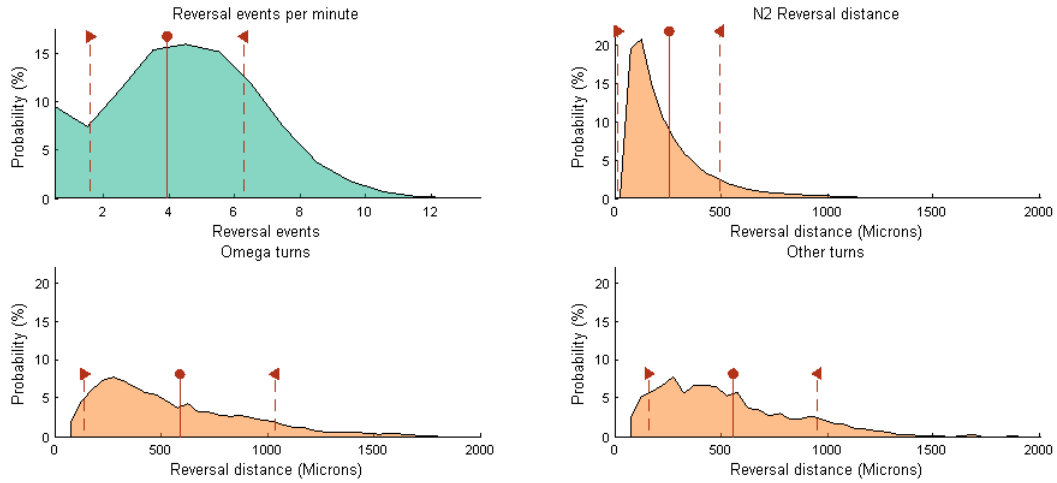


Figure 4.7: Wild type (N2) reversal rate and distance. Characteristics based on 82040 reversal events for wild type strain were extracted. A rate of 3.9±2.3 (±SD) reversal events per minute was identified for N2 data. Bin size of 1 was used for the histogram. The distance mean of reversals was 255±238.5 (±SD) microns (μ). For 4397 omega turns correlated with reversals a mean of 586.5±446.3 μ (±SD) was found, for 1629 other turns a mean of 555.3±396.3 μ (±SD) was identified. Bin size of 50 μ was used for distance histograms.

lated with reversals - 77±1.9%. The turns that were not correlated with reversals in omega bends amounted to only 3±0.5%, this number was larger (13±1.5%) for other turns.

Reversals are an integral part of navigation behaviour and we have also studied the frequency and distance of reversals for the dataset of 82040 reversal events in wild type (N2) data (Figure 4.7). The mean rate of reversals was 3.9±2.3 (±SD) events per minute. The mean distance of reversals correlated with omega bends and other turns was found to be larger than the mean distance of all of the reversals that the animal makes. A mean reversal distance for wild type strain is 255±238.5 (±SD) microns and 586.5±446.3 μ (±SD) and 555.3±396.3 μ (±SD) for omega and other turns respectively. It shows that escape response characterized by larger reversals are more likely to take place when the animals are turning. This data can be used in simulations of navigation behaviour because high sample numbers allow having a confident representation of the probability distribution for each of the reversal features.

4.2.5 Phenotypic turn analysis of *C. elegans* mutants

Turn correlation to reversals, turn ventral bias and overall reversal rate was computed for 350 strains for both omega and other turns. In addition, we have added a long reversal measure for both omega and other turns. Reversals longer than one body length of the animal and correlated to turns were also identified. These eight measures were compared to N2 data and helped identify significantly different strains. Statistical significance was computed using Chi-Square test with Bonferroni correction and error bars for turn counts were computed using bootstrapping with 95% confidence interval (Figure 4.9). A large list of strains can be identified that have significantly different turn characteristics. A general observation can be made from Figure 4.9 - mutations that have significant difference from the N2 lie on one side of the turning feature probability. In Figure 4.9 A all significant strains have a smaller probability of turn correlation with reversals. Turns are less likely to be correlated with reversals in these strains. In Figure 4.9 B all significant strains have higher probability - the length of reversals is less tightly controlled in the significant strains (it has a low probability for N2). Similarly in Figure 4.9 C almost all significantly different strains have lower probability of ventral turns. The only exceptions are four wild isolates (JU345, JU393, JU402, ED3054) and goa-1 (sa734, DG1856) mutant. Figure 4.9 C also indicates loss of a tightly controlled turn characteristic - ventral bias. Significantly different reversal mutants show less reversals and more extra long reversals. These findings indicate that turning behaviour in wild type (N2) animals is tightly controlled and genetic perturbations cause loss of these tendencies across the turning features that we have compared.

In order to identify phenotypes for the chosen turn characteristics heatmaps were generated for p-values obtained using Chi-Square test with Bonferroni correction. The values were binned to fall into intervals between 0, 0.0005, 0.005, 0.05 (from very significant to not significant). Dark orange indicates probabilities higher than that of the N2 and dark green indicates probabilities smaller to that of the N2 while white squares indicate no statistical significance. Track length, forward velocity and backward velocity feature comparisons to the wild type (N2) were also provided to help identify more interesting strains that have phenotypes in turning characteristics but otherwise have normal locomotion. Three groups of strains were extracted - strains having significantly different turn and reversal

correlation probabilities to that of N2, Figure 4.10. Strains having significantly different turn and long reversal correlation probabilities, Figure 4.11. In addition, strain with significantly different ventral bias, Figure 4.12 and lastly significantly different reversal rate from the wild type (N2), Figure 4.13. Note that there is some overlap between the figures. In addition to the newly defined phenotypes in navigation behaviour there were some several strains that stood out.

Besides uncoordinated (UNC) mutants in Figure 4.10 there is a whole list of interesting strains that show no phenotype in locomotion speed but show a pronounced loss of turn and reversal correlation. Loss of correlation with reversals in both omega and other turns can be seen in loss of function mec-12 (e1605) mutant which is one of a dozen genes required for touch receptor neuron function in *C.elegans* [73]. Loss of function egl-19 (n2368) shows no locomotion velocity phenotype but has a phenotype in turn and reversal correlation. Egl-19 plays a pivotal role in regulating muscle excitation and contraction [74]. Other strains with the phenotype are displayed in Figure 4.10.

Two genes acd-5 (ok2657) (here referred to as T28F2.7) and asic-2 (ok289) were previously hypothesised to share similar function or even operate in the same channel complex in some cells [43]. They both are degenerin/epithelial Na^+ (DEG/ENaC) channels and both showed two behavioural motifs that are different from the N2 - a bout in forward locomotion and a pause in a curved body shape. Both mutants were further from the forward locomotion motif and closer to the curved pause motif. The mutants could be distinguished from the N2 but not from each other in context of these two motifs [43].

T28F2.7 (ok2657) and asic-2 (ok289) come up significant in both turn reversal correlation loss Figure 4.10 and ventral bias loss Figure 4.12 indicating a possibility that previously detected motifs were related to different navigation behaviour. Loss of ventral bias means that the animals made more dorsal turns and their shape would look different (coiled dorsally) to that of N2 (coiled ventrally).

Reversal phenotype was identified in egl-47 (n1081), a gene that encodes two orphan G-protein-coupled receptor isoforms, which share all seven transmembrane domains but have different extracellular N termini [75]. Loss of function mutation shows no detectable egg-laying defects but gain of function mutations and overexpression studies suggest that egl-47 is required for negatively regulating activity of the hermaphrodite-specific motor neurons (HSNs) that control egg laying. GFP expression shows egl-47 localization in a small subset of head

neurons, the HSNs (the only cells in egg laying system in adults) and the PVQ tail interneurons [75]. Our data shows a phenotype in reversal rate as well as long reversal correlation with other turns Figure 4.13. Reversal rate phenotype in egl-47 suggests that there is a navigation phenotype which could be further studied to characterize genes function.

Other navigation phenotypes can be found in Figures 4.10, 4.11, 4.12, 4.13.

4.3 Discussion and Conclusion

We have developed a novel turn detection algorithm that allowed unbiased and more biologically inspired identification of reorientation events. Instead of identifying specific body shapes we looked for rapid changes in the direction of locomotion. We have analysed 327.7 hours of wild type data and identified 6416 turning events.

In addition to turns containing omega bends we have discovered turns that have other posture characteristics - wide bend where the animal doesn't bring head close to the body and pivoting behaviour that involves reversal and head bend to facilitate a turn. Reorientation angle for omega and other turns shows similar characteristics indicating importance of other turns in navigation behaviour.

Wild type (N2) animals have high turn and reversal correlation and high ventral bias. Our results here show the same properties and allows a better estimation of the probabilities due to the fact that to our knowledge we have used the largest N2 turn dataset to date. In addition, other turn characteristics like reversal length, duration, rate and posture before the turn open up a wealth of data. It allows better understanding of the dynamics of reorientation and is a very valuable resource for computational modelling. Using the probabilities described here for turn characteristics as well as reversal rate and distance better computational models can be build and the dynamics of the navigation behaviour studied in more detail.

In addition to characterizing N2 turning behaviour we have also extracted turning behaviour from the data of 350 mutants presented in the online *C.elegans* behavioural database. We compared four turn characteristic probabilities - turn and reversal correlation, turn and long reversal correlation, ventral bias and re-

versal rate - of each of these mutants to the N2. We have observed the loss of turn and reversal correlation and ventral bias as well as increase in reversal length when turning which indicates that wild type (N2) turning behaviour is a tightly controlled process and loss of function mutations disturb this process. We have identified turn phenotypes for many uncoordinated (unc) mutants as well as strains with a defect in genes with broad molecular effects. Specifically we were able to identify previously discussed (DEG/ENaC) channel mutants acd-5 (ok2657) (here referred to as T28F2.7) and asic-2 (ok289). Their association was hypothesized by eigen-worm analysis which identified specific behavioural motifs that two strains shared [43]. The results presented here show an increase in dorsal bias which accounts for preciously identified unique shapes.

While clustering approaches can help identify broad categories and analyse pathways, studying specific behavioural events such as turning can help to clearly identify behaviours contributing to predicted phenotypic similarity and allow generating specific hypotheses about functional gene interactions. Combining machine learning and statistical methods to cluster high dimensional feature data and approaches similar to the one presented here can be a very effective way to generate accurate hypothesis and efficiently guide experimental work.

It is important to note that while utmost care has been taken to accurately extract navigation phenotypes from the mutant dataset the strains have not been outcrossed in a systematic manner due to time constraints. There is a chance that spontaneous mutations have appeared in strains that show phenotypes and further experimental work must be done to make sure none of additional mutations that could cause the observed phenotype exist.

Results presented here showcase the potential of online behavioural dataset available at *www.wormbehavior.mrc-lmb.cam.ac.uk*. All of the analysis has been done using the feature data files that are available to download from our online resource. Identifying important behavioural events that have not been addressed by the initial analysis can continue generating knowledge and become a rich resource for exploratory analysis of *C. elegans* behaviour.

4.4 Acknowledgements

I would like to thank Andrè Brown, Buyun Zhao and Barry Bentley for fruitful discussion and support during this work.

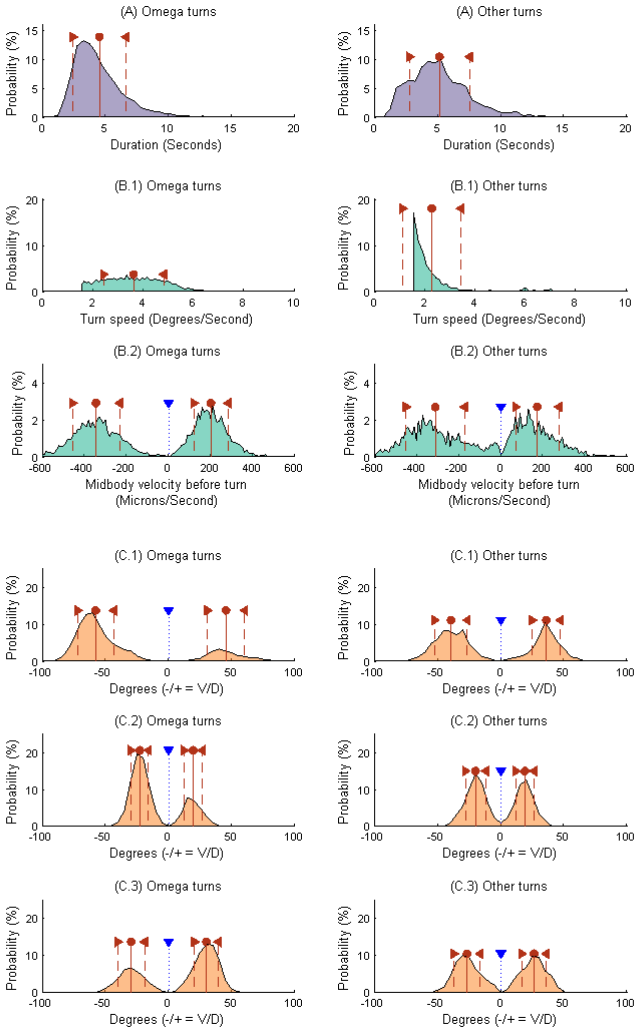


Figure 4.8: Wild type (N2) turn characteristics. Turn analysis based on 4543 omega turns and 1873 other turns. **A**. *Turn duration.* Bin size of 0.5 seconds was used. Omega mean 4.5 ± 2 sec (± SD), other turns mean 5.2 ± 2.4 sec (± SD). **B.1.** *Turn rate.* Bin size for the histograms was 0.1 degrees/second (deg/sec). Omega turn mean 3.6 ± 1.2 deg/sec (± SD), other turn mean 2.3 ± 1.1 deg/sec (± SD). **B.2.** *Maximum velocity.* Maximum forward or backward velocity one second before the turn was identified. Bin size of 10 microns/second (μ/sec) was used. Omega turn backward mean 345 ± 112 μ/sec (± SD) and forward mean of 201 ± 81 μ/sec (± SD). For other turns a mean of 311 ± 138 μ/sec (± SD) for backward and 172 ± 102 μ/sec (± SD). **C.** *Curvature of head, midbody and tail.* Measured as maximum curvature during one second before the turn. Bin size of 5° was used for the histograms. **C.1-3** *Head.* Omega turns ventral mean 58° ± 14° (± SD), dorsal mean 39° ± 13° (± SD). Other turns: ventral mean 39° ± 13° (± SD), dorsal mean 36° ± 11° (± SD). *Midbody.* Omega turns ventral mean 23° ± 7° (± SD), dorsal mean 19° ± 7° (± SD). Other turns ventral mean 20° ± 8° (± SD), dorsal mean 19° ± 7° (± SD). *Tail.* Omega turn dorsal bends had a mean of 30° ± 9° (± SD). Ventral mean was 30° ± 11° (± SD). Other turns had almost identical distributions and probabilities for dorsal and ventral angles - a mean of 26° ± 10° (± SD).

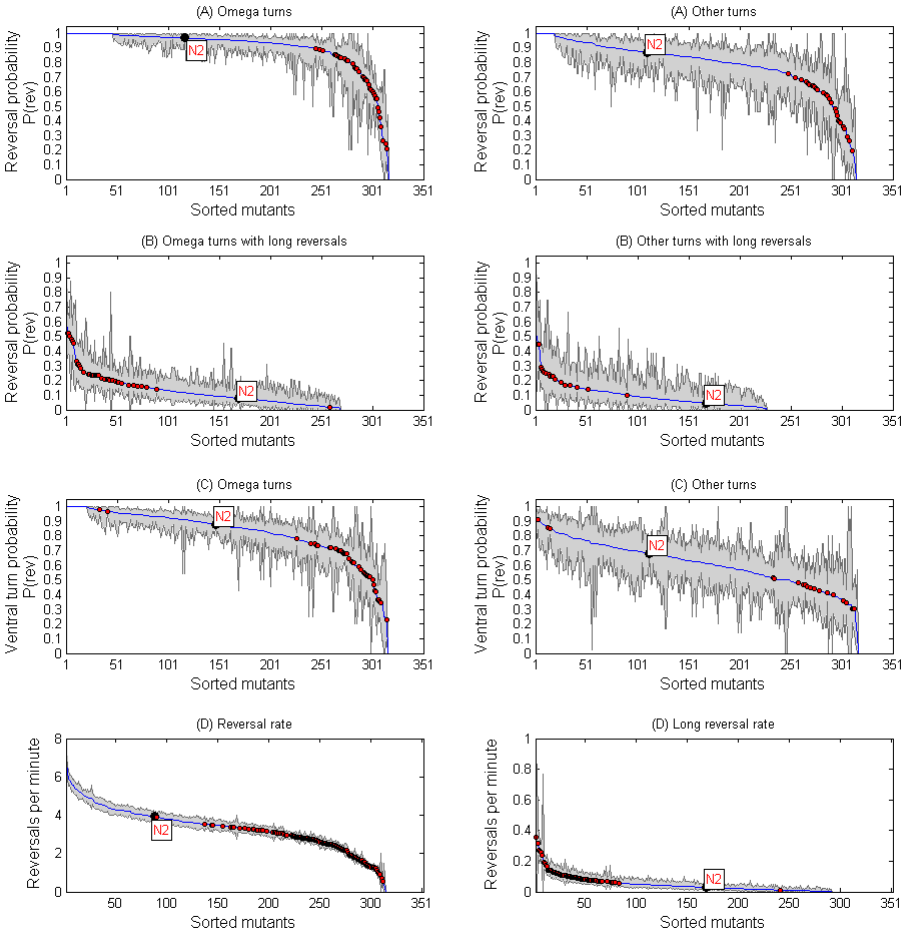


Figure 4.9: Phenotyping overview. Strain and wild type (N2) comparison. Red dots signify significantly different strains using Chi-Square test, [N2] denotes wild type, error bars are computed using bootstrapping with 90% confidence interval. The strains at the bottom are sorted according to the probability magnitude. **A**. Turn and reversal correlation probability for omega and other turns. **B**. Turn and long reversal correlation probability for omega and other turns. **C**. Probability of ventral turns for omega and other turns. **D**. Overall reversal rate per minute comparison.

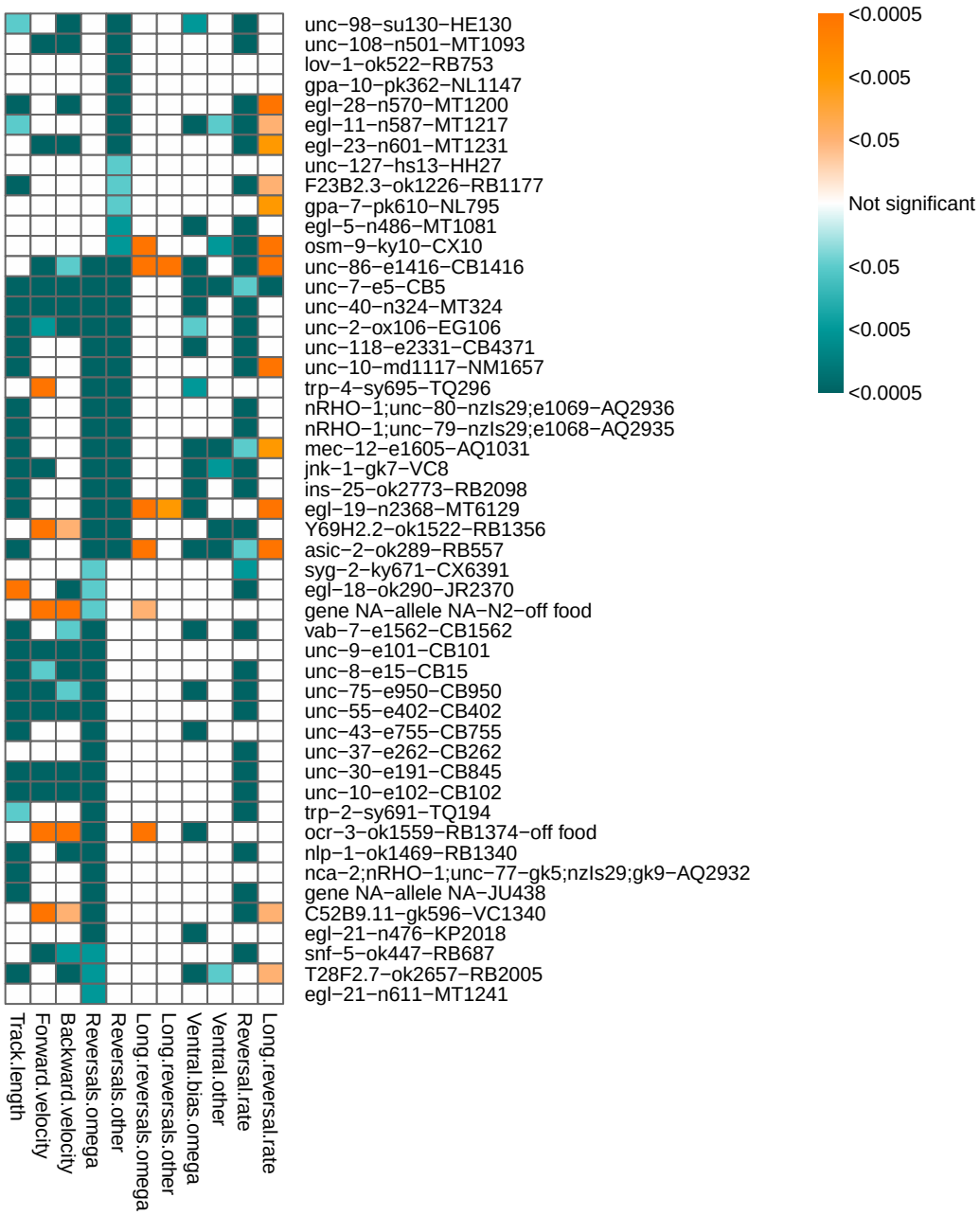


Figure 4.10: Strains with phenotype in turn and reversal correlation. Strains with significantly different turn and reversal probabilities have been extracted. A total of 49 strains were identified, out of witch 22 strains were found for omega turns, 15 strains were found for both omega and other turns and 12 strains were found for other turns. All significant strains had lower turn and reversal correlation probability.

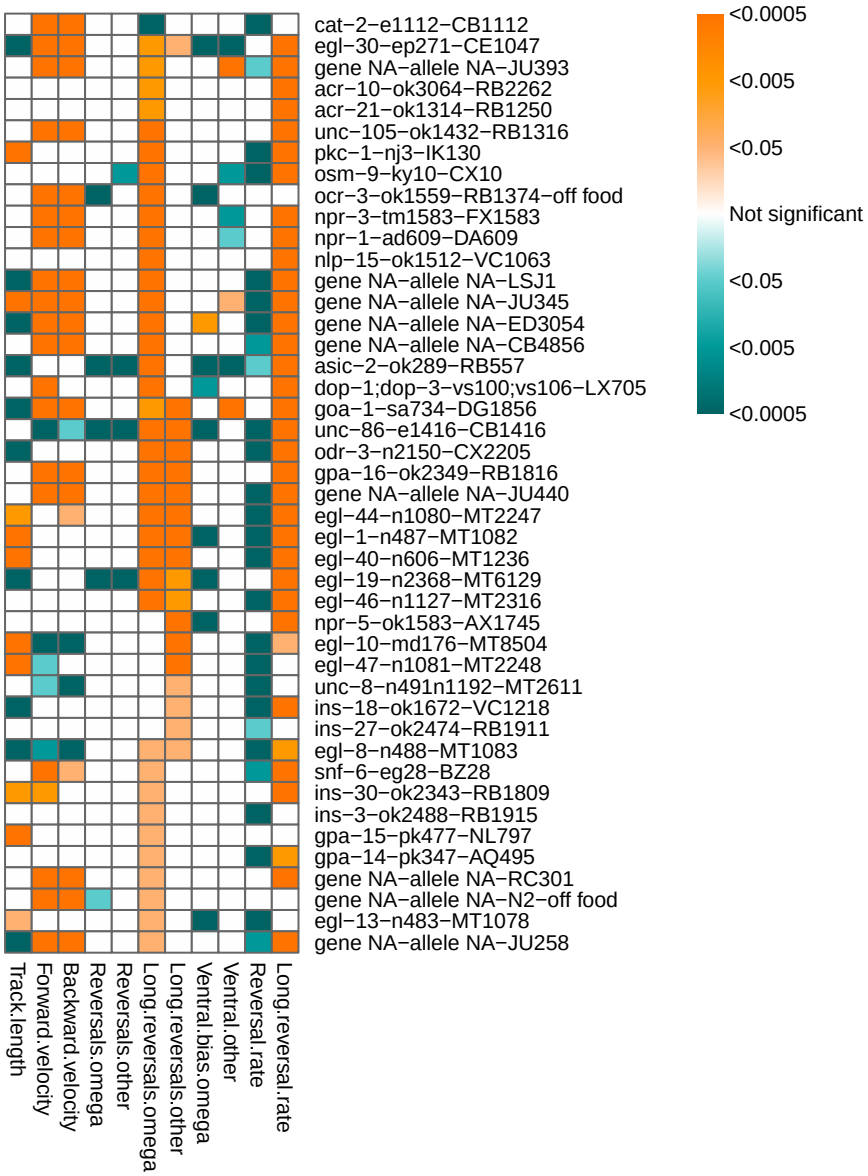


Figure 4.11: Strains with a phenotype in turn correlation with long reversals. Strains with significantly different turn and long reversal probabilities have been extracted. A total of 44 strains were identified, out of witch 26 strains were found for omega turns, 12 strains were found for both omega and other turns and 6 strains were found for other turns. All significant strains had higher turn and long reversal correlation probability except cat-2 (e1112, CB1112) which had lower probability.

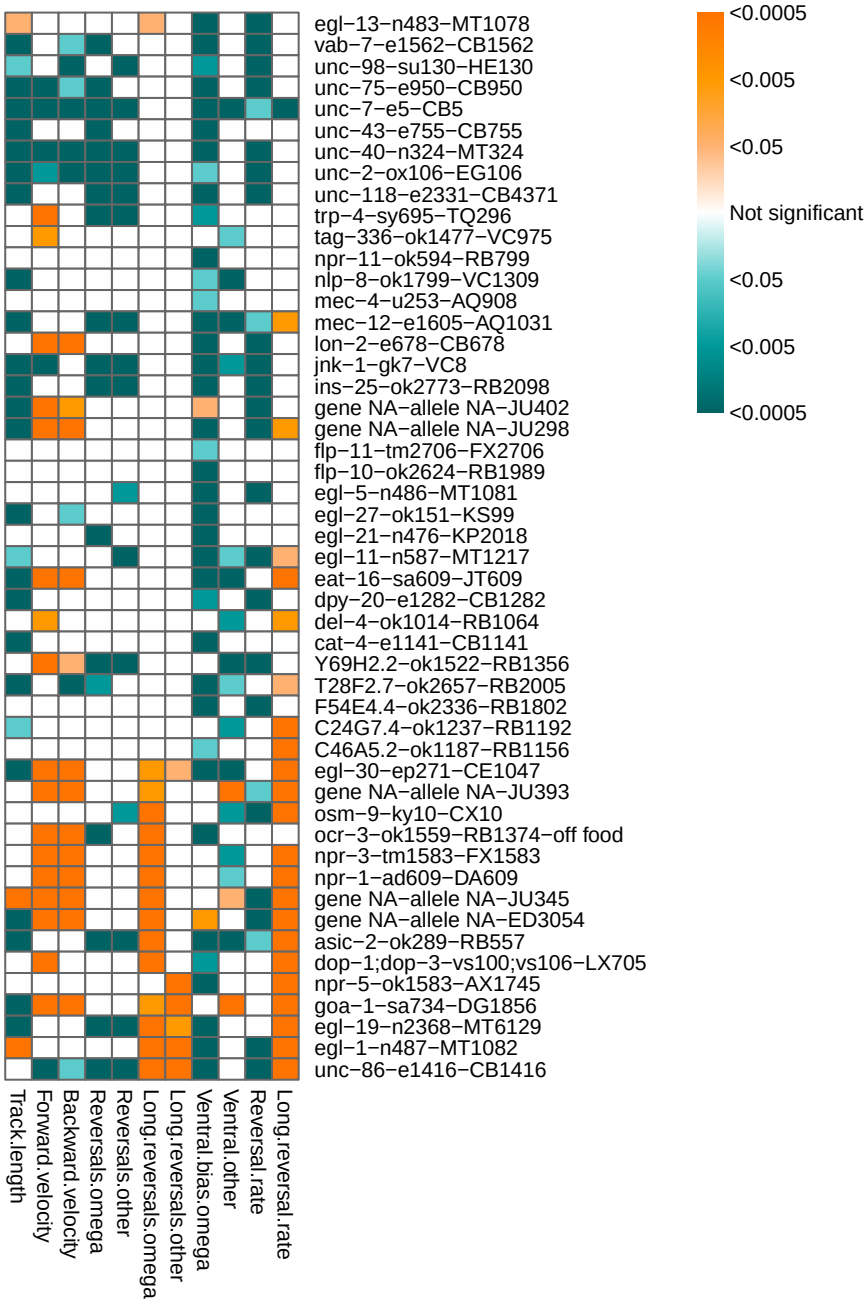


Figure 4.12: Strains with phenotype in turn ventral bias. Strains with significantly different turn ventral bias are displayed. A total of 50 strains were identified, out of witch 31 strains were found for omega turns, 9 strains were found for both omega and other turns and 10 strains were found for other turns. All significant strains had lower turn ventral bias except four wild isolates (JU345, JU393, JU402, ED3054) and goa-1 (sa734, DG1856).

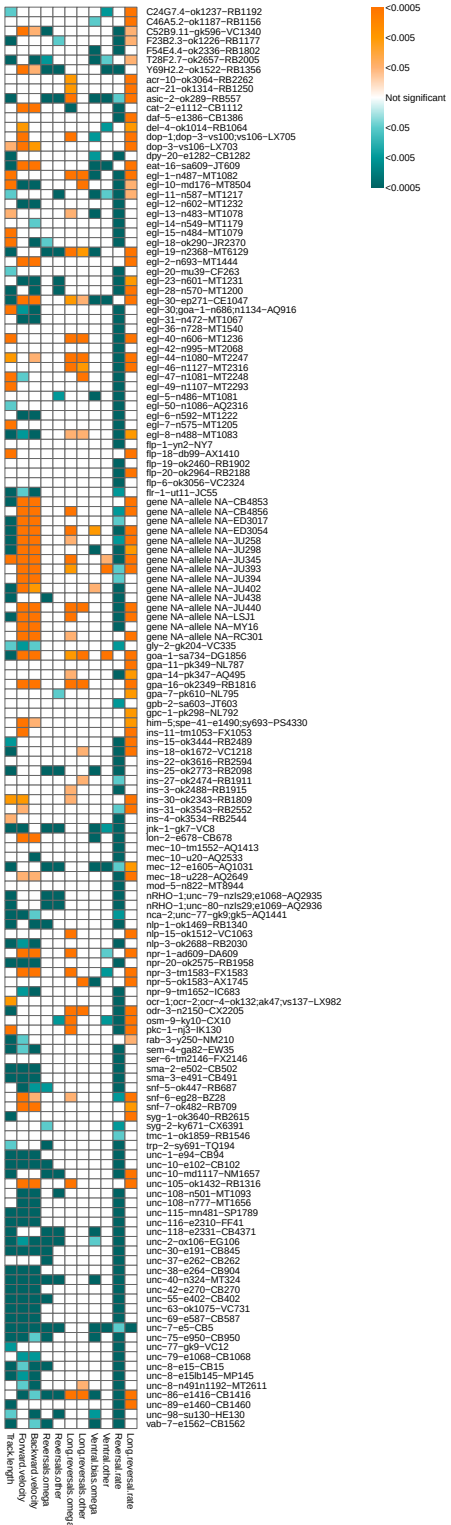


Figure 4.13: Strains with phenotype in reversal rate. Strains with significantly different reversal rates were identified. A total of 148 strains were found, out of witch 81 strains were found for omega turns, 36 strains were found for both omega and other turns and 31 strains were found for other turns. All significant strains had lower reversal rate for omega bends and higher reversal rate for other turns.

CHAPTER 5

OUTLOOK

The platform described in this thesis is a continuing project that has undergone many improvements and additions over the course of its development. It has reached a functional steady-state and can rapidly deliver results as described in this thesis and related publications. However, there are several immediate improvements that could greatly benefit the platform and make it an even more attractive tool for *C. elegans* scientific community. Several of these future projects are described below.

5.1 High resolution multi-worm tracking

Digital video camera technology has undergone a rapid development resulting in an increase in supported resolution, frame rate and data throughput as well as a rapid decrease in it's cost. One of the companies in the forefront of novel digital camera technology is a Canadian company called *Point Grey*. They offer affordable and highly innovative cameras using USB 3.0 framework and capable of up to 9 mega pixel (MP) resolution and recording of 150 frames per second (FPS) [76]. Such device would allow multi worm tracking with the resolution of a single worm tracker. It would allow collecting data of a large number of worms much quicker without sacrificing the detail at which the shape of the animals is characterized. Data acquired and analysed with such a system would be complementary to the data presented in the online *C. elegans* behavioural database.

We have acquired one of their cameras and through the joint efforts of André Brown and Barry Bentley we have been able to use the existing segmentation and analysis algorithms developed for our current platform to analyse the data acquired with the new high resolution camera. Even though a large volume of work still needs to be done it is clear that such approach would be able to directly interface with existing software and open up a path to a platform that would not require motorized stages, would be easier to build and enable much higher throughput.

5.2 Protocol for external data submission

Online *C. elegans* behavioural database is a static repository serving the data acquired by the Schafer group to the scientific community. A number of laboratories have acquired and successfully used the tracker platform presented in this thesis. However, currently there is no way for the data generated by other laboratories to be compared to and included in the online database. Since the tracking protocols and algorithms are the same a simple set of steps would allow other researchers to compare their data to the online database and upload it for the use by other research groups.

According to the tracking protocol each of the experiments has a set of wild type animals tracked for normalization and background control purposes. Comparing the wild type data acquired in another laboratory to the rich wild type dataset collected by the Schafer group would allow determining if the experiments done in that laboratory are compatible with the data presented in the online database. In case the wild type data falls within expected variance, behavioural phenotypes of the new data could be identified and the whole dataset included in the online database. Access to this data could be restricted to suit the needs of contributing researchers and subsequently opened up to other research groups.

The increasing number of laboratories using our proposed platform would ensure development of standardized protocols and algorithms for *C. elegans* behaviour analysis. It would also grow the online database and ensure that it does not rely on one group's efforts. Rich and vibrant repository for *C. elegans* behavioural data would allow unprecedented scope and accuracy of analysis.

5.3 Beyond spontaneous behaviour

The data presented in this thesis describes spontaneous *C. elegans* behaviour. However, the model organism is capable of much more. The exact extent to which the nervous system of *C. elegans* could be probed quantitatively is not exactly known. In order to explore it further, behavioural observations beyond spontaneous behaviour will have to be performed.

It is known that *C. elegans* responds to a wide array of external stimuli. Assays of gentle touch show animals reversing when touched on their anterior half. In contrast, when touched on their posterior half the animals accelerate and move forward. The gentle touch responses have been described in detail and have been shown to be controlled by specific neural circuits [63]. *C. elegans* nematodes also sense soluble and volatile chemicals such as biotin, cAMP, salt, carbon dioxide and oxygen as well as respond to changes in the temperature gradient [77], [78].

It has been shown that *C. elegans* animals rapidly adapt to changes to the viscosity of the environment suggesting a delicate integration of senses that could be studied further [79], [80], [81], [82]. In addition, further work in combining genetic and neuronal ablations can help characterize the nervous system and enable better theoretical models [83], [84], [85], [86].

Novel methods based on microfluidic devices could provide the required tools for precise animal immobilization and stimulation. A review by Nikos Chronis on microfluidic applications for *C. elegans* details many of the possibilities [87]. In addition, optogenetics based methods to apply controlled stimuli directly to *C. elegans* neurons can also be used. A review by Fenno *et. al.* can be referred to for more details [88].

A relatively little amount of alterations to the algorithms and the pipeline presented in this thesis would enable studies of *C. elegans* nervous system beyond those based on the spontaneous behaviour. Such studies would use the full extend of the versatility offered by this remarkable model organism and progress the field to new and exciting areas.

Bibliography

[1] Antoine Barriere and Marie-Anne Felix. High local genetic diversity and low outcrossing rate in caenorhabditis elegans natural populations. *Current biology : CB*, 15(13):1176–1184, July 2005.

[2] Harksun Lee, Myung-kyu Choi, Daehan Lee, Hye-sung Kim, Hyejin Hwang, Heekyeong Kim, Sungsu Park, Young-ki Paik, and Junho Lee. Nictation, a dispersal behavior of the nematode caenorhabditis elegans, is regulated by IL2 neurons. *Nat Neurosci*, 15(1):107–112, January 2012.

[3] C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, 282(5396):2012–2018, December 1998.

[4] K Yook, TW Harris, T Bieri, A Cabunoc, J Chan, WJ Chen, P Davis, N de la Cruz, A Duong, R Fang, U Ganesan, C Grove, K Howe, S Kadam, R Kishore, R Lee, Y Li, HM Muller, C Nakamura, B Nash, P Ozersky, M Paulini, D Raciti, A Rangarajan, G Schindelman, X Shi, EM Schwarz, M Ann Tuli, K Van Auken, D Wang, X Wang, G Williams, J Hodgkin, M Berriman, R Durbin, P Kersey, J Spieth, L Stein, and PW Sternberg. WormBase 2012: more genomes, more data, new website. *Nucleic acids research*, 40(Database issue):D735–41, January 2012.

[5] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. System of the nematode caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 314(1165):1–340, 1986.

[6] S Brenner. The genetics of caenorhabditis elegans. *Genetics*, 77(1):71–94, May 1974.

[7] J. Ahringer. Reverse genetics (April 6, 2006). *WormBook, ed. The C. elegans Research Community, WormBook*, pages 1551 – 8507, April 2006.

[8] AG Fraser, RS Kamath, P Zipperlen, M Martinez-Campos, M Sohrmann, and J Ahringer. Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature*, 408(6810):325–330, November 2000.

[9] RS Kamath, AG Fraser, Y Dong, G Poulin, R Durbin, M Gotta, A Kanapin, N Le Bot, S Moreno, M Sohrmann, DP Welchman, P Zipperlen, and J Ahringer. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, 421(6920):231–237, January 2003.

[10] F Simmer, M Tijsterman, S Parrish, SP Koushika, ML Nonet, A Fire, J Ahringer, and RH Plasterk. Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. *Current biology : CB*, 12(15):1317–1319, August 2002.

[11] B Lehner, A Calixto, C Crombie, J Tischler, A Fortunato, M Chalfie, and AG Fraser. Loss of LIN-35, the Caenorhabditis elegans ortholog of the tumor suppressor p105Rb, results in enhanced RNA interference. *Genome biology*, 7(1):R4, 2006.

[12] S Kennedy, D Wang, and G Ruvkun. A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature*, 427(6975):645–649, February 2004.

[13] NA Croll. Behavioural analysis of nematode movement. *Advances in parasitology*, 13:71–122, 1975.

[14] R. Gaugler and A.L. Bilgrami. *Nematode Behaviour*. CABI Publishing Series. CABI Pub., 2004.

[15] Katie S Kindt, Veena Viswanath, Lindsey Macpherson, Kathleen Quast, Hongzhen Hu, Ardem Patapoutian, and William R Schafer. Caenorhabditis elegans TRPA-1 functions in mechanosensation. *Nat Neurosci*, 10(5):568–577, May 2007.

[16] David B. Dunsenbery. Video camera-computer tracking of nematode-caenorhabditis elegans to record behavioral responses. *Journal of Chemical Ecology*, 11(9):1239–1247, 1985.

[17] David B. Dusenbery. Using a microcomputer and video camera to simultaneously track 25 animals. *Computers in biology and medicine*, 15(4):169–175, 1985.

[18] Phillip L. Williams and David B. Dusenbery. Aquatic toxicity testing using the nematode, Caenorhabditis elegans. *Environmental Toxicology and Chemistry*, 9(10):1285–1290, 1990.

[19] R Dhawan, DB Dusenbery, and PL Williams. Comparison of lethality, reproduction, and behavior as toxicological endpoints in the nematode Caenorhabditis elegans. *Journal of toxicology and environmental health. Part A*, 58(7):451–462, December 1999.

[20] M de Bono and CI Bargmann. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. *Cell*, 94(5):679–689, September 1998.

[21] DR Soll. The use of computers in understanding how animal cells crawl. *International review of cytology*, 163:43–104, 1995.

[22] Jonathan T. Pierce-Shimomura, Thomas M. Morse, and Shawn R. Lockery. The fundamental role of pirouettes in caenorhabditis elegans chemotaxis. *The Journal of Neuroscience*, 19(21):9557–9569, 1999.

[23] LA Hardaker, E Singer, R Kerr, G Zhou, and WR Schafer. Serotonin modulates locomotory behavior and coordinates egg-laying and movement in Caenorhabditis elegans. *Journal of neurobiology*, 49(4):303–313, December 2001.

[24] LE Waggoner, GT Zhou, RW Schafer, and WR Schafer. Control of alternative behavioral states by serotonin in Caenorhabditis elegans. *Neuron*, 21(1):203–214, July 1998.

[25] Z Feng, CJ Cronin, JH Wittig, PW Sternberg, and WR Schafer. An imaging system for standardized quantitative analysis of C. elegans behavior. *BMC bioinformatics*, 5:115, August 2004.

[26] Christopher Cronin, Jane Mendel, Saleem Mukhtar, Young-Mee Kim, Robert Stirbl, Jehoshua Bruck, and Paul Sternberg. An automated system

for measuring parameters of nematode sinusoidal movement. *BMC Genetics*, 6(1):5, 2005.

[27] GD Tsibidis and N Tavernarakis. Nemo: a computational tool for analyzing nematode locomotion. *BMC neuroscience*, 8:86, 2007.

[28] D Ramot, BE Johnson, TL Berry, L Carnell, and MB Goodman. The Parallel Worm Tracker: a platform for measuring average speed and drug-induced paralysis in nematodes. *PloS one*, 3(5):e2208, 2008.

[29] Nicholas A Swierczek, Andrew C Giles, Catharine H Rankin, and Rex A Kerr. High-throughput behavioral analysis in c. elegans. *Nature Methods*, 8(7):592–598, July 2011.

[30] G Tsechpenakis, L Bianchi, D Metaxas, and M Driscoll. A novel computational approach for simultaneous tracking and feature extraction of C. elegans populations in fluid environments. *IEEE transactions on bio-medical engineering*, 55(5):1539–1549, May 2008.

[31] SD Buckingham and DB Sattelle. Fast, automated measurement of nematode swimming (thrashing) without morphometry. *BMC neuroscience*, 10:84, 2009.

[32] SH Simonetta and DA Golombek. An automated tracking system for Caenorhabditis elegans locomotor behavior and circadian studies application. *Journal of neuroscience methods*, 161(2):273–280, April 2007.

[33] DS Matthies, PA Fleming, DM Wilkes, and RD Blakely. The Caenorhabditis elegans choline transporter CHO-1 sustains acetylcholine synthesis and motor function in an activity-dependent manner. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(23):6200–6212, June 2006.

[34] Andrew M Leifer, Christopher Fang-Yen, Marc Gershow, Mark J Alkema, and Aravinthan D T Samuel. Optogenetic manipulation of neural activity in freely moving Caenorhabditis elegans. *Nat Meth*, 8(2):147–152, February 2011.

[35] JN Stirman, M Brauner, A Gottschalk, and H Lu. High-throughput study of synaptic transmission at the neuromuscular junction enabled by optogenetics and microfluidics. *Journal of neuroscience methods*, 191(1):90–93, August 2010.

[36] S Faumont, G Rondeau, TR Thiele, KJ Lawton, KE McCormick, M Sottile, O Griesbeck, ES Heckscher, WM Roberts, CQ Doe, and SR Lockery. An image-free opto-mechanical system for creating virtual environments and imaging neuronal activity in freely moving caenorhabditis elegans. *PloS one*, 6(9):e24666, 2011.

[37] Steven J. Husson, Wagner Steuer Costa, Cornelia Schmitt, and Alexander Gottschalk. Keeping track of worm trackers. *WormBook, ed. The C. elegans Research Community, WormBook*, September 2010.

[38] Kuang-Man Huang, Pamela Cosman, and William R. Schafer. Automated detection and analysis of foraging behavior in caenorhabditis elegans. *Journal of Neuroscience Methods*, 171(1):153 – 164, 2008.

[39] Raphael Sznitman, Manaswi Gupta, Gregory D. Hager, Paulo E. Arratia, and Josué Sznitman. Multi-Environment Model Estimation for Motility Analysis of Caenorhabditis elegans. *PLoS ONE*, 5(7):e11631, 07 2010.

[40] E. Fontaine, J. Burdick, and A. Barr. *Automated Tracking of Multiple C. Elegans*. 2006.

[41] N Roussel, CA Morton, FP Finger, and B Roysam. A computational model for C. elegans locomotory behavior: application to multiworm tracking. *IEEE transactions on bio-medical engineering*, 54(10):1786–1797, October 2007.

[42] Greg J. Stephens, Bethany Johnson-Kerner, William Bialek, and William S. Ryu. Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol*, 4(4):e1000028, 04 2008.

[43] Andre E. X. Brown, Eviatar I. Yemini, Laura J. Grundy, Tadas Jucikas, and William R. Schafer. A dictionary of behavioral motifs reveals clusters of genes affecting caenorhabditis elegans locomotion. *Proceedings of the National Academy of Sciences*, 110(2):791–796, 2013.

[44] Synametrics Technologies. DeltaCopy - Rsync for Windows. *http://www.aboutmyip.com/AboutMyXApp/DeltaCopy.jsp, Accessed: 2013-07-01.*

[45] Eviatar Yemini, Tadas Jucikas, Laura J Grundy, Andre E X Brown, and William R Schafer. A database of caenorhabditis elegans behavioral phenotypes. *Nat Meth*, advance online publication, July 2013.

[46] Arun K. Ramani, Tungalag Chuluunbaatar, Adrian J. Verster, Hong Na, Victoria Vu, Nadge Pelte, Nattha Wannissorn, Alan Jiao, and Andrew G. Fraser. The majority of animal genes are required for wild-type fitness. *Cell*, 148(4):792 – 802, 2012.

[47] David Houle, Diddahally R. Govindaraju, and Stig Omholt. Phenomics: the next challenge. *Nature Reviews Genetics*, 11(12):855–866, December 2010.

[48] Anne C. Hart. Behavior. *WormBook, ed. The C. elegans Research Community, WormBook*, July 2006.

[49] W. Geng, P. Cosman, C.C. Berry, Zhaoyang Feng, and W.R. Schafer. Automatic tracking, feature extraction and classification of c. elegans phenotypes. *Biomedical Engineering, IEEE Transactions on*, 51(10):1811–1820, 2004.

[50] Kuang-Man Huang, Pamela Cosman, and William R. Schafer. Machine vision based detection of omega bends and reversals in c. elegans. *Journal of Neuroscience Methods*, 158(2):323 – 336, 2006.

[51] John D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[52] Todd Starich, Ji Xu, I Martha Skerrett, Bruce Nicholson, and Jocelyn Shaw. Interactions between innexins UNC-7 and UNC-9 mediate electrical synapse specificity in the caenorhabditis elegans locomotory nervous system. *Neural Development*, 4(1):16, 2009.

[53] Edward Yeh, Sharon Ng, Mi Zhang, Magali Bouhours, Ying Wang, Min Wang, Wesley Hung, Kyota Aoyagi, Katya Melnik-Martinez, Michelle Li, Fang Liu, William R Schafer, and Mei Zhen. A Putative Cation Channel,

NCA-1, and a Novel Protein, UNC-80, Transmit Neuronal Activity in C. elegans. *PLoS Biol*, 6(3):e55, 03 2008.

[54] Thomas Boulin, Marc Gielen, Janet E. Richmond, Daniel C. Williams, Pierre Paoletti, and Jean-Louis Bessereau. Eight genes are required for functional reconstitution of the caenorhabditis elegans levamisole-sensitive acetylcholine receptor. *Proceedings of the National Academy of Sciences*, 105(47):18590–18595, 2008.

[55] Antony M. Jose, I. Amy Bany, Daniel L. Chase, and Michael R. Koelle. A specific subset of transient receptor potential vanilloid-type channel subunits in caenorhabditis elegans endocrine cells function as mixed heteromers to promote neurotransmitter release. *Genetics*, 175(1):93–105, 2007.

[56] Zhaoyang Feng, Wei Li, Alex Ward, Beverly J. Piggott, Erin R. Larkspur, Paul W. Sternberg, and X.Z. Shawn Xu. A C. elegans model of nicotine-dependent behavior: regulation by TRP-Family channels. *Cell*, 127(3):621–633, November 2006.

[57] Rajarshi Ghosh, Aylia Mohammadi, Leonid Kruglyak, and William Ryu. Multiparameter behavioral profiling reveals distinct thermal response regimes in caenorhabditis elegans. *BMC Biology*, 10(1):85, 2012.

[58] Eviatar Yemini, Rex A. Kerr, and William R. Schafer. Preparation of samples for single-worm tracking. *Cold Spring Harbor Protocols*, 2011(12):pdb.prot066993, 2011.

[59] N. Otsu. A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions*, 9(1):62–66, 1979.

[60] Herbert Freeman. On the encoding of arbitrary geometric configurations. *Electronic Computers, IRE Transactions on*, EC-10(2):260–268, 1961.

[61] J.E. Sulston and H.R. Horvitz. Post-embryonic cell lineages of the nematode, caenorhabditis elegans. *Developmental Biology*, 56(1):110 – 156, 1977.

[62] Mark J. Alkema, Melissa Hunter-Ensor, Niels Ringstad, and H. Robert Horvitz. Tyramine functions independently of octopamine in the caenorhabditis elegans nervous system. *Neuron*, 46(2):247–260, April 2005.

[63] M. Chalfie, J. E. Sulston, J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The neural circuit for touch sensitivity in caenorhabditis elegans. *The Journal of Neuroscience*, 5(4):956–964, 1985.

[64] Jesse M. Gray, Joseph J. Hill, and Cornelia I. Bargmann. A circuit for navigation in caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3184–3191, 2005.

[65] William S. Ryu and Aravinthan D. T. Samuel. Thermotaxis in caenorhabditis elegans analyzed by measuring responses to defined thermal stimuli. *The Journal of Neuroscience*, 22(13):5727–5733, 2002.

[66] Hatim A. Zariwala, Adam C. Miller, Serge Faumont, and Shawn R. Lockery. Step response analysis of thermotaxis in caenorhabditis elegans. *The Journal of Neuroscience*, 23(10):4369–4377, 2003.

[67] Neil A. Croll. Indolealkylamines in the coordination of nematode behavioral activities. *Canadian Journal of Zoology*, 53(7):894–903, 1975. PMID: 1079747.

[68] Neil A. Croll and James M. Smith. Integrated behaviour in the feeding phase of caenorhabditis elegans (nematoda). *Journal of Zoology*, 184(4):507–517, 1978.

[69] Jonathan T. Pierce-Shimomura, Michael Dores, and Shawn R. Lockery. Analysis of the effects of turning bias on chemotaxis in c. elegans. *Journal of Experimental Biology*, 208(24):4727–4733, 2005.

[70] Erwin Kreyszig. Introductory mathematical statistics : Principles and methods. *John Wiley*, page 206, 1970.

[71] T.J. DiCicio and B. Efron. Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228, 1996.

[72] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Monographs on statistics and applied probabilities. Chapman & Hall/CRC, 1993.

[73] T. Fukushige, Z.K. Siddiqui, M. Chou, J.G. Culotti, C.B. Gogonea, S.S. Siddiqui, and M. Hamelin. Mec-12, an alpha-tubulin required for touch sensitivity in c. elegans. *Journal of Cell Science*, 112(3):395–403, 1999.

[74] Raymond Y.N. Lee, Leslie Lobel, Michael Hengartner, H.Robert Horvitz, and Leon Avery. Mutations in the [alpha]1 subunit of an l-type voltageactivated ca2+ channel cause myotonia in caenorhabditis elegans. *EMBO J*, 16(20):6066–6076, October 1997.

[75] James J. Moresco and Michael R. Koelle. Activation of egl-47, a gao-coupled receptor, inhibits function of hermaphrodite-specific motor neurons to regulate caenorhabditis elegans egg-laying behavior. *The Journal of Neuroscience*, 24(39):8522–8530, 2004.

[76] Point Grey Research. Reliable end-to-end imaging over USB 3.0. *http://ww2.ptgrey.com/USB3, Accessed: 2013-08-11.*

[77] C. Bargmann. Chemosensation in C. elegans (October 25, 2006). *WormBook, ed. The C. elegans Research Community, WormBook*, October 2006.

[78] L Luo, DA Clark, D Biron, L Mahadevan, and AD Samuel. Sensorimotor control during isothermal tracking in Caenorhabditis elegans. *The Journal of experimental biology*, 209(Pt 23):4652–4662, December 2006.

[79] A Vidal-Gadea, S Topper, L Young, A Crisp, L Kressin, E Elbel, T Maples, M Brauner, K Erbguth, A Axelrod, A Gottschalk, D Siegel, and JT Pierce-Shimomura. Caenorhabditis elegans selects distinct crawling and swimming gaits via dopamine and serotonin. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42):17504–17509, October 2011.

[80] Jeremie Korta, Damon A. Clark, Christopher V. Gabel, L. Mahadevan, and Aravinthan D. T. Samuel. Mechanosensation and mechanical load modulate the locomotory gait of swimming C. elegans. *Journal of Experimental Biology*, 210(13):2383–2389, 2007.

[81] F Lebois, P Sauvage, C Py, O Cardoso, B Ladoux, P Hersen, and JM Di Meglio. Locomotion control of Caenorhabditis elegans through confinement. *Biophysical journal*, 102(12):2791–2798, June 2012.

[82] JH Boyle, S Berri, M Tassieri, IA Hope, and N Cohen. Gait modulation in C. elegans: It's not a choice, it's a reflex! *Frontiers in behavioral neuroscience*, 5:10, 2011.

[83] C Fang-Yen, M Wyart, J Xie, R Kawai, T Kodger, S Chen, Q Wen, and AD Samuel. Biomechanical analysis of gait adaptation in the nematode Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47):20323–20328, November 2010.

[84] J Sznitman, PK Purohit, P Krajacic, T Lamitina, and PE Arratia. Material properties of Caenorhabditis elegans swimming at low Reynolds number. *Biophysical journal*, 98(4):617–626, February 2010.

[85] P Sauvage, M Argentina, J Drappier, T Senden, J Simon, and JM Di Meglio. An elasto-hydrodynamical model of friction for the locomotion of Caenorhabditis elegans. *Journal of biomechanics*, 44(6):1117–1122, April 2011.

[86] R. Mailler, J. Avery, J. Graves, and N. Willy. *A biologically accurate 3D model of the locomotion of caenorhabditis elegans*. 2010.

[87] N Chronis, M Zimmer, and CI Bargmann. Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans. *Nature methods*, 4(9):727–731, September 2007.

[88] L Fenno, O Yizhar, and K Deisseroth. The development and application of optogenetics. *Annual review of neuroscience*, 34:389–412, 2011.

[89] Chen-An Tsai and James J. Chen. Multivariate analysis of variance test for gene set analysis. *Bioinformatics*, 25(7):897–903, 2009.

[90] Beibei Zhao, Parul Khare, Lisa Feldman, and Joseph A. Dent. Reversal frequency in caenorhabditis elegans represents an integrated response to the state of the animal and its environment. *The Journal of Neuroscience*, 23(12):5319–5328, 2003.

[91] Patrick T. McGrath, Matthew V. Rockman, Manuel Zimmer, Heeun Jang, Evan Z. Macosko, Leonid Kruglyak, and Cornelia I. Bargmann. Quantitative mapping of a digenic behavioral trait implicates globin variation in c. elegans sensory behaviors. *Neuron*, 61(5):692–699, March 2009.

[92] Katherine P. Weber, Subhajyoti De, Iwanka Kozarewa, Daniel J. Turner, M. Madan Babu, and Mario de Bono. Whole genome sequencing highlights

genetic changes associated with laboratory domestication of C. elegans. *PLoS ONE*, 5(11):e13922, 11 2010.

APPENDIX A

Chapter 2 Appendix

A.1 MySQL database table list

A full list of MySQL database tables used in the *C. elegans* behavioural experiment analysis pipeline.

Table A.1: Database tables. Tables comprising experiment analysis and annotation database. In addition to table name and description primary keys and foreign keys are specified.

Table name	**Primary Key**	**Foreign Key**	**Desscription**
experiments	id	-	Main table containing all unique experiments.
segmentation-ExperimentList	id	id	Analysis table containing a list of all experiments that need to be analysed, their locations and their analysis status (columns *taken* and *completed*).

exp_annotation	id	strainID, alleleID, geneID, trackerID, sexID, ageID, ventralSideID, agarSideID, foodID, habitID, locationID, experimenterID, genotypeID, treatment	Table containing annotations for each experiment.
experimenters	expID	-	A list of experimenters that carried out tracking experiments.
experiments_experimenters	id	expID, locationID	Table listing experiments tracked by each experimenter.
experimenterLocation	id	locationID	Table listing location where experiments were carried out.
strain	strainID	-	Table containing all strain names.
allele	alleleID	-	Table containing all allele names.
gene	geneID	-	Table containing all gene names.
chromosome	chromosomeID	-	Table containing all chromosome names.
sex	sexID	-	Table containing possible sex for experimental animals.
age	ageID	-	Table containing possible age of experimental animals.

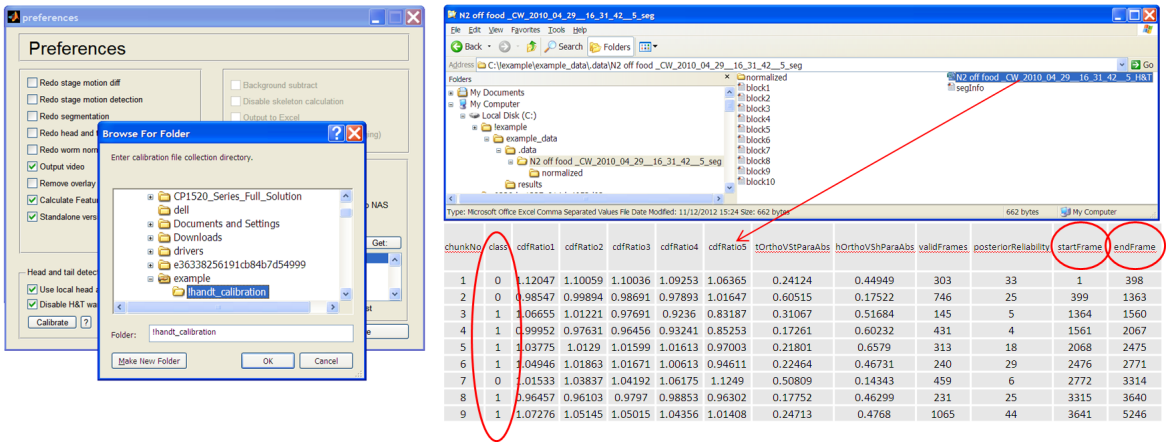
wormSide	sideID	-	Table containing possible side on which the animals crawled during the experiments.
food	foodID	-	Table containing possible food conditions of the experiments.
habituation	habitID	-	Table containing possible habituation conditions of the experiments.
trackerNo	trackerID	-	Table containing all tracker names.
genotype	genotypeID	cgcID	Table containing all tracker names.
ventralSide	ventralSideID	-	Table containing annotation of the ventral side of the animal.
treatment	treatmentID	-	Table containing possible chemical treatments during the experiments.
trackedStrains	id	strainID, alleleID, geneID, chromosomeID, genotypeID	List of all the tracked strains.
cgcDB	cgcID	-	Table containing strain list, genotype and annotations. Obtained from CGC website.
version	id	id	Records version of the analysis used for each step for analysed experiments.

frame_warning_labels	entryid	id	Frame warning annotation table. Each entry contains possible frame warning message.
errors	id	id	A table containing errors that occurred during the analysis run for each of the experiments.
aviFileSize	id	id	Table recording video file size for each experiment.
freeHDDSpace	computerID	compu-terID	Table to help keep track of remaining free space on tracker computers.
youtube	id	id	Table containing unique YouTube keys for each experiment.
issuuid	strainID	strainID	Unique ISSUU ID for each strain.

A.2 Head and tail detection calibration

To calibrate automatic head and tail detection users are instructed to create an empty directory that will contain head and tail assignment files and form a training data set. Head and tail assignment files are produced at a segmentation step and are stored in an analysis directory that is created by the toolbox (the directory is called *.data*). There is one file produced for each experiment. Example path for the annotation file is: *C:/example/.data/N2 off food_CW_2010_04_29__16_31_42__5_seg/N2 off food_CW_2010_04_29__16_31_42__5 _H&T.csv*). More head and tail assignment files will ensure better training set and will increase the predictive power of the algorithm.

For each of the head and tail assignment files users have to manually edit the *class* column, Figure A.1. It can be done by opening video overlay file and keeping the assignments unchanged if head and tail is identified correctly and switching



chunkNo	class	cdfRatio1	cdfRatio2	cdfRatio3	cdfRatio4	cdfRatio5	tOrthoVStParaAbs	hOrthoVShParaAbs	validFrames	posteriorReliability	startFrame	endFrame
1	0	1.12047	1.10059	1.10036	1.09253	1.06365	0.24124	0.44949	303	33	1	398
2	0	0.98547	0.99894	0.98691	0.97893	1.01647	0.60515	0.17522	746	25	399	1363
3	1	1.06655	1.01221	0.97691	0.9236	0.83187	0.31067	0.51684	145	5	1364	1560
4	1	0.99952	0.97631	0.96456	0.93241	0.85253	0.17261	0.60232	431	4	1561	2067
5	1	1.03775	1.0129	1.01599	1.01613	0.97003	0.21801	0.6579	313	18	2068	2475
6	1	1.04946	1.01863	1.01671	1.00613	0.94611	0.22464	0.46731	240	29	2476	2771
7	0	1.01533	1.03837	1.04192	1.06175	1.1249	0.50809	0.14343	459	6	2772	3314
8	1	0.96457	0.96103	0.9797	0.98853	0.96302	0.17752	0.46299	231	25	3315	3640
9	1	1.07276	1.05145	1.05015	1.04356	1.01408	0.24713	0.4768	1065	44	3641	5246

Figure A.1: Automatic head and tail detection calibration. **Left** is a preferences directory selection window. **Top right**, location of head and tail assignment file that is created after the first pass of analysis. **Bottom right**, example of head and tail assignment file indicating continuous head and tail frame sequences and their assignment based on the motion and coloration statistics.

it (from 0 to 1 or the other way around) in case the overlay shows a wrong assignment. Continuous head and tail sequences are identified and *startFrame* and *endFrame* columns can be used to rapidly check the assignment correctness. Head and tail assignment for frames between *startFrame* and *endFrame* will always be the same.

Automatic head and tail assignment calibration algorithm will load all of the assignment files, their head and tail assignment annotation and corresponding statistics for each of the continuous video segments. It will use the data to create a training set that will be used by the toolbox for head and tail detection.

APPENDIX B

CHAPTER 3 APPENDIX

Declaration. *For the sake of completeness this appendix contains some figures and work that has been carried out by Andrè Brown and Eviatar Yemini.*

B.1 Testing strain and group significance

We used Wilcoxon rank-sum to test the differences between each group's features and the wild type (329 groups of strains and time-based N2 groups by 702 measures), and we used Fisher's exact test for measurements found exclusively in the experiment or control groups. Because all features were measured using automated algorithms, scoring was effectively blind. We controlled false discovery rate across these comparisons, converting P values to their q equivalents [51]. Each group was assigned its minimum q value as a measure of group significance. Under this paradigm, every group was found to be significantly different from the wild type at $q = 0.05$. To ensure this was not an artefact of our methodology, we chose a second method to verify our result. As sparse sampling led to a difficulty in assessing normality, and as dimensionality outnumbered observations for all but the lab N2 data set, we measured the Hotelling T2 statistic using a shrinkage estimation of the covariance matrix and a permutation test to determine P values [89]. In four strains, at least one measure was detected exclusively in either the strain or its control, meaning that the measurement was always observed in one set and never in the other (e.g., some strains never perform reversals). When this occurred, we used a Fisher's exact test to measure the probability that our sets

were drawn from the same distribution of observed and unobserved events. For all other groups, we ran 10,001 permutations. We corrected the P values, controlling the false discovery rate, to their q equivalents. Once again, every group was found to be significantly different from the wild type at $q = 0.05$, bearing out the results of our primary methodology.

B.2 Reproducibility and sensitivity of measurements

We performed several tests to assess the sensitivity and reproducibility of the measurements in our database. To assess variability within an individual animal, we tracked 25 young-adult wild-type hermaphrodites for 2 h without the usual habituation period we perform for our regular assays. Consistent with previous reports [56], [90], the worm speed was well fit by an exponential decay, with a time constant $t = 19$ min ($R2 = 0.96$). Because speed might be expected to correlate with both crawling amplitude and frequency, we investigated whether these two measures might show similar habituation. In fact, crawling frequency (titled 'Absolute Midbody Frequency' in our measurements) was best fit with an exponential decay of $t = 19$ min ($R2 = 0.71$; Appendix B Figure B.1 B), whereas amplitude ('Absolute Midbody Amplitude') decayed with a longer time constant of $t = 48$ min ($R2 = 0.59$; Appendix B Figure B.1 A). Thus, the time necessary for locomotion features to adapt to steady state appears to differ, even among features affecting similar aspects of behaviour.

B.3 Wild type (N2) variability

We next assessed the between-animal variability of worms from the same wild-type stock. Individuals of a given stock, after generations of self-fertilization, would be expected to be nearly genetically identical; however, small differences in age as well as difficult-to-control environmental factors might be expected to affect behavioural data. We recorded a total of $1,218$ N2 young-adult hermaphrodites over the course of 3 years, from 2009 to 2012. Animals were tracked from January to December, from Tuesday through Friday, from 9 a.m. to 5 p.m. We chose six representative measures to analyze variability: length, forward speed ('Positive

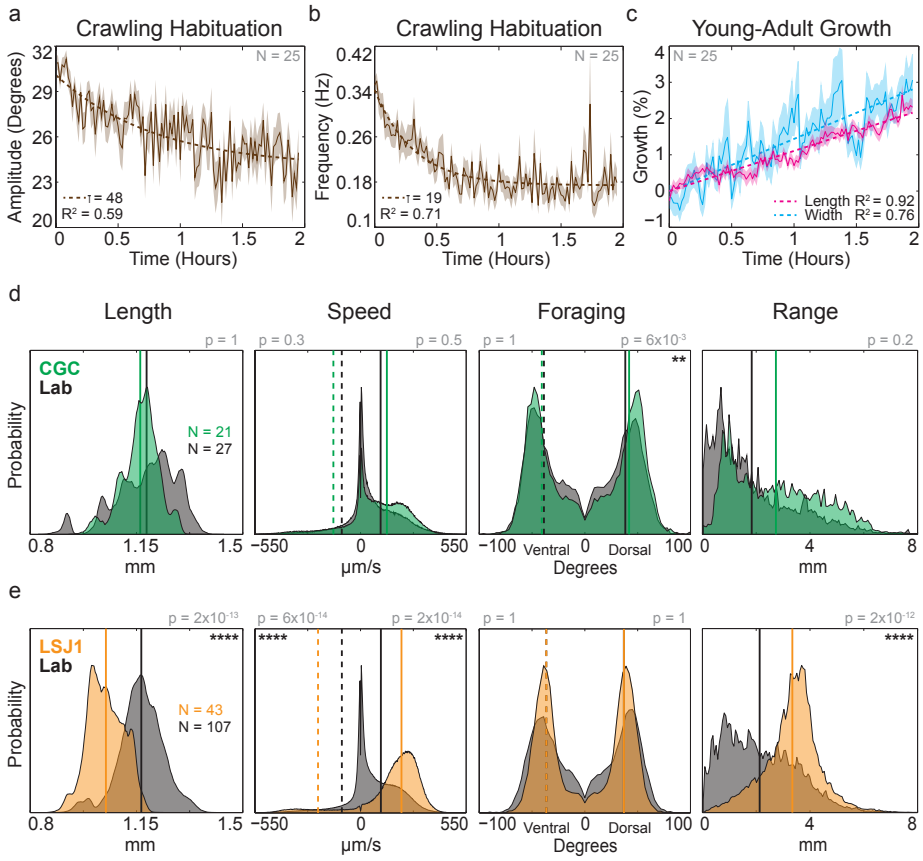


Figure B.1: Phenotypic summaries using selected features for subsets of strains. Colours in the heat maps are used to indicate the q value for each feature for the comparison between each of the mutant strains listed on the right and the N2 reference data. Red values indicate features that have a significantly higher value in the mutant, whereas blue indicates significantly lower values in the mutant. Genes and features were both hierarchically clustered for easier comparison. **A**. Subset of mutant strains with previously known locomotion phenotypes. **B**. Subset of mutant strains with no previously reported locomotion phenotype. Abs, absolute; max, maximum.

Midbody Speed'), foraging amplitude ('Absolute Foraging Amplitude'), reversals ('Backward Motion Frequency'), coiling frequency and exploratory range (Appendix B Figure B.1 D and Figure B.2). To assess the influence of these factors (hour, day and month), we performed one-way ANOVAs with Bonferroni correction for 18 tests and checked for significance at $P = 0.05$. The tested groups showed a mixture of normality and non-normality when we used a Shapiro-Wilk test with correction for multiple comparisons. Therefore, to avoid the assumption of normality, we also performed Kruskal-Wallis tests with the same correction and

a, which resulted in identical significance. Because all tracked animals were identified as fourth-stage larvae the night before, we reasoned that animals tracked later in the day should be slightly older. Indeed, there was a small but significant difference in length (and other features) among animals tracked throughout the day. We measured changes in length and midbody width, over 2 h, among the previously mentioned group of 25 young-adult, lab-stock N2s and found them well fit with a 1% linear growth per hour: $R2 = 0.92$ and 0.76, respectively. This may explain the observed hourly differences among animals. Likewise, we reasoned that although we control our lab to maintain it at 22 °C, the temperature may still vary slightly with the season, whereas the day of the week should have no consistent trend. Indeed, the day of the week had no significant effect on any of the tested parameters, whereas the month of the year had a slight but significant effect on all parameters except coiling (Appendix B Figure B.2). So that seasonal effects were accounted for, all data were controlled by wild types collected within a 2-week window centred around the experiments.

With the variability in mind, we computed the number of worms required to achieve statistical power in discriminating phenotypic differences. We bootstrapped Wilcoxon rank-sum tests, comparing multiple group sizes. The tested groups were chosen from our three-year collection of wild-type lab stock and compared to ones chosen from an identical collection, mean shifted by the discriminatory amount. The findings indicate that 10 worms provide well over 90% power to discriminate 2 s.d. of mean difference. Moreover, 20 worms discriminate a single s.d. at over 80% power (Appendix B Figure B.2). For example, when testing forward speed, 20 worms discriminate a mean difference of at least 53 μm/s with 90% power (assuming the compared groups share an identically shaped distribution to our lab-stock N2).

B.4 Assessing variation between different wild-type stocks

To estimate genetically derived variability between different lab stocks of the same strain, we compared the behaviour of our lab N2 stock to one obtained from the Caenorhabditis Genetics Centre. We chose four common measures (Appendix B Figure B.1 D) to assess potential differences: length, speed (‘Midbody

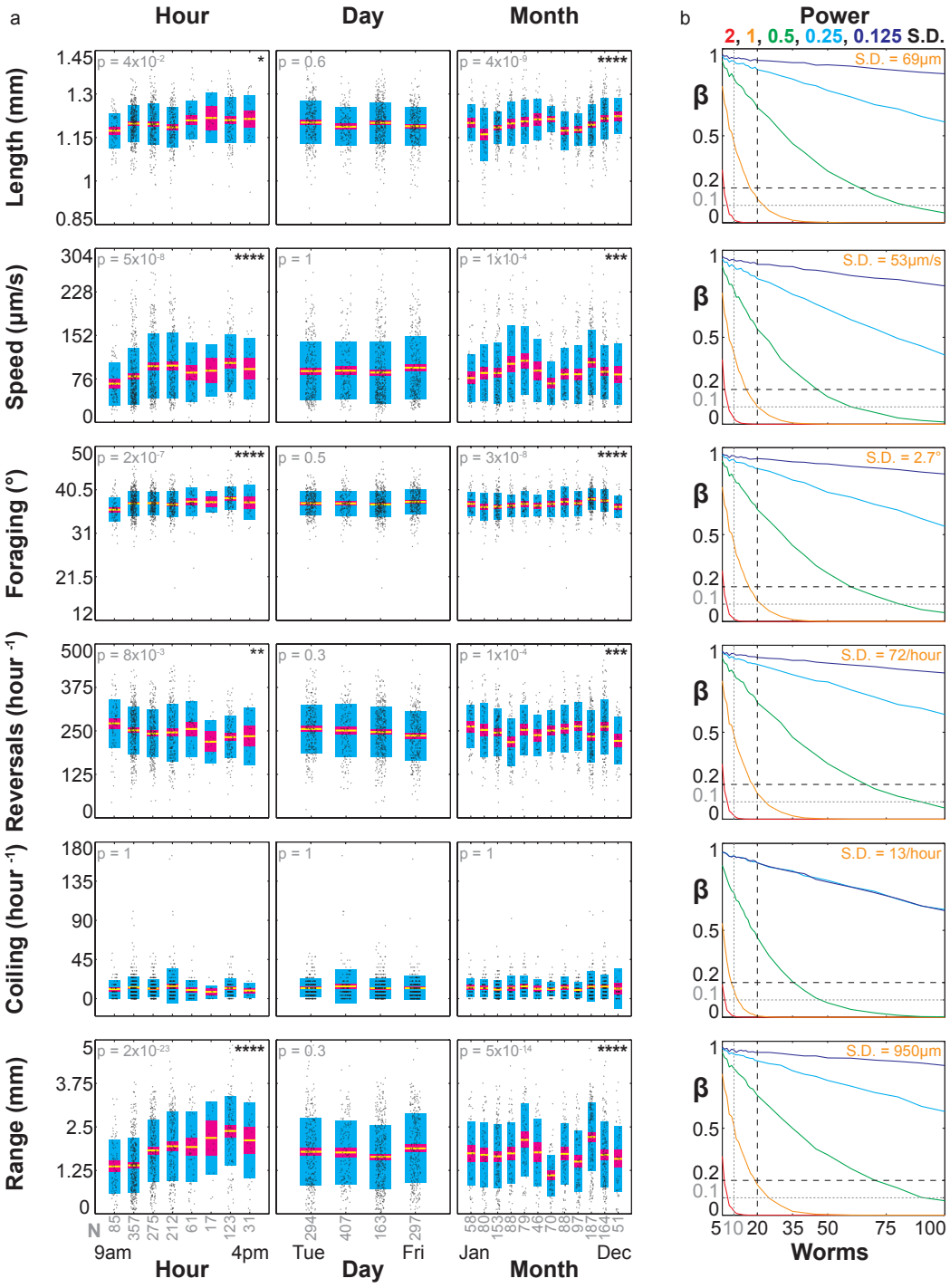


Figure B.2: Next page.

Speed'), foraging amplitude and range. Speed and foraging were further subdivided into forward-backward and dorsal-ventral statistics, yielding a total of six statistical comparisons. Shapiro-Wilk tests, with $a = 0.05$, indicated a mix-

Figure B.2: Wild type variability. (Previous page). 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 $p = 0.05$, 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 x 10^{-2} *, day = 0.6, month = 4 x 10^{-9} ****), forward speed (*p-value* hour = 5 x 10^{-8} ****, day = 1, month = 1 x 10^{-4}, ***), foraging amplitude ignoring the dorsal-ventral orientation (*p-value* hour = 2 x 10^{-7} ****, day = 0.5, month = 3 x 10^{-8} ****), reversal frequency (*p-value* hour = 8 x 10^{-3} **, day = 0.3, month = 1 x 10^{-4} ***), coiling frequency (*p-value* hour = 1, day = 1, month = 1), and exploratory range (*p-value* hour = 2 x 10^{-23} ****, day = 0.3, month = 5 x 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**. β (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. β 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.

ture of normal and non-normal distributions for our 21 *CGC* and 27 lab N2 measures. Therefore, we compared both groups using Wilcoxon rank-sum tests with Bonferroni correction. Of all these comparisons, only dorsal-oriented foraging ($P = 6$ x 10^{-3}) showed a significant difference between the two wild-type stocks, although visually the histograms for other features showed some apparent differences; further testing, with subdivided feature measurements, uncovered related significant measures (available online in our database). Nonetheless, many features appeared relatively consistent between the two wild-type stocks despite likely genetic divergence.

We also analysed *LSJ1*, a more distant relative of N2 (Appendix B Figure B.1 E). This strain is descended from the same wild isolate as N2, but the two strains have diverged following years of laboratory cultivation [91], [92]. We compared 43 *LSJ1* with 107 of our lab N2 for the same six features described above. As expected, the LSJ1 showed significantly higher speed (backward $P = 2$ x 10^{-14};

forward $P = 6$ x 10^{-14}) and a significantly larger range of exploration ($P = 2$ x 10^{-12}), results correlating with the bordering phenotype of *LSJ1*. Additionally, *LSJ1* was slightly (0.1 mm) but significantly shorter in length ($P = 2$ x 10^{-13}) than N2. In comparison, the summary statistics for foraging are nearly overlapping ($P = 1$ for both sides) with a dorsal mean and s.e.m. of 37° ± 0.5° and 36.9° ± 0.4° for the *LSJ1* and lab N2, respectively, and ventral values of 36.5° ± 0.3° for both strains.

B.5 Algorithm details

B.5.1 Morphology Features

Length. Worm length is computed from the segmented skeleton by converting the chain-code pixel length to microns.

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.

Area. The worm area is computed from the number of pixels within the segmented contour. The sum of the pixels is converted to microns^2.

Area/Length.

Midbody Width/Length.

B.5.2 Posture Features

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. 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 chaincode length, in opposing directions, along the skeleton. When insufficient skeleton points are present, the angle re-

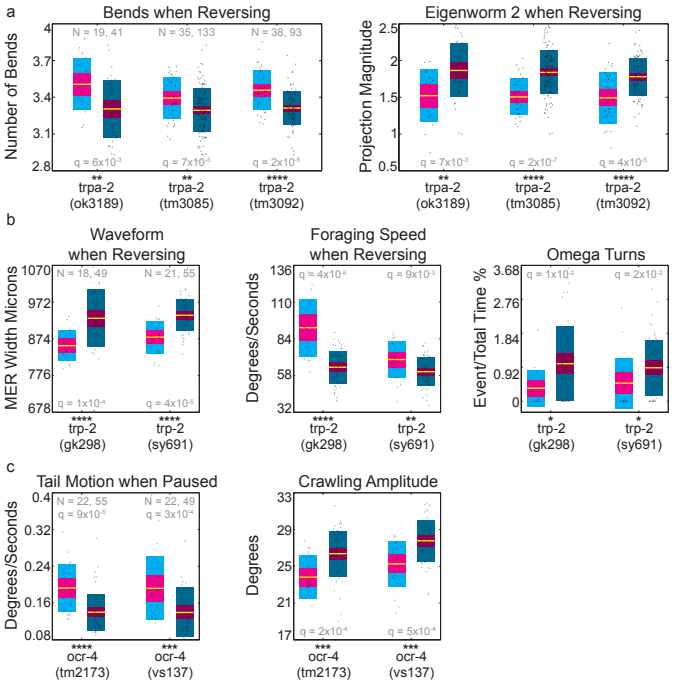


Figure B.3: New locomotion phenotypes for three TRP channels. New locomotion phenotypes for three TRP channels. Both *trpa-2* and *ocr-4* have no characterized phenotypes despite previous experimental publications. We used Wilcoxon rank-sum tests and False-Discovery Rate (FDR) to determine *q-values* for each comparison. Corrections were performed for 329 groups (strains and time-based collections of N2s) by 702 features. Significance was set at $q = 0.05$. 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 ($q = 6 \text{ x } 10^{-3}$, $7 \text{ x } 10^{-3}$, $2 \text{ x } 10^{-5}$). In further support, all three alleles show a decrease in their second eigenworm projection when reversing ($q = 7 \text{ x } 10^{-3}$, $2 \text{ x } 10^{-7}$, $4 \text{ x } 10^{-5}$), a 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 ($q = 1 \text{ x } 10^{-4}$, $4 \text{ x } 10^{-5}$). Their foraging speed, when reversing, is faster ($q = 4 \text{ x } 10^{-6}$, $9 \text{ x } 10^{-3}$). 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 ($q = 1 \text{ x } 10^{-2}$, $2 \text{ x } 10^{-2}$). **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 ($q = 9 \text{ x } 10^{-5}$, $3 \text{ x } 10^{-3}$). Furthermore, both *ocr-4* alleles display smaller amplitudes when crawling ($q = 2 \text{ x } 10^{-3}$, $5 \text{ x } 10^{-3}$).

mains 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.

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 a 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.

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).

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' [26]. 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.

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' [26]. 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.

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' [26].

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.

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 [42]. 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 re-

maining 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.

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).

B.5.3 Motion Features

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). 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 anticlockwise, 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.

Motion States. The worm's forward, backward, and paused motion states attempt to differentiate these event states unambiguously (Figure 3.5 F). 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 labelling 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).

Crawling. Worm crawling is expressed as both an amplitude and frequency (Figure 3.5 E). 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 centre 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, centred 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.

Foraging. Worm foraging is expressed as both an amplitude and an angular speed (Figure 3.5 G). 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 6Hz7 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 a of 2.5 and normalized such that the filter integrates to 1. The foraging amplitude is defined as the largest foraging angle measured, prior to crossing 0°. In other words, the largest nose bend prior to returning to a straight, unbent position. Therefore, the foraging amplitude time series follows a discrete, stair-step pattern. The amplitude is signed negatively whenever the nose points towards the worm's ventral side. The foraging angular speed is measured as the foraging angle difference between subsequent frames divided by the time between these frames. To centre 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.

Turns. Omega and upsilon turn events are computed similarly to a previously described method9 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° at the first but not the last third of the body (the head but not the tail). The middle third must then exceed 30°. And finally, the last but not the first third of the body must exceed 30° (the tail but not the head). This sequence of a 30° mean supplementary angle, passing continuously along the worm from head to tail, is labeled an omega turn event. Upsilon turns are computed nearly identically but they capture all events that escaped being labeled omega turns, wherein the mean supplementary angle exceeded 15° on one side of the worm (the first or last third of the body) while not exceeding 30° on the opposite end.

B.5.4 Path Features

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 (Figure 3.5 H).

Dwelling. The worm dwelling is computed for the head, midbody, tail, and the entire worm (Figure 3.5 I). 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.

Curvature. The path curvature is defined as the angle, in radians, of the worm's path divided by the distance it travelled 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 travelled, 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.

B.6 Feature File Overview

The features are presented within four types of files available online at:
www.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 labelled with a mean and standard deviation of 95.17 $\pm$ 17.5% across individual videos and 95.69% of the frames collectively.

Outliers can compress visual details in their corresponding histograms. For

this reason, the strain collections underwent one more filtering step prior to inclusion in the PDF files. Experiments were discarded wherein any of the worm data exceeded reasonable bounds of 250 to 2000 microns for length, 25 to 250 microns for width, and/or -1000 to 1000 microns/seconds for the midbody speed. Outliers were seldom found. Overall, 49 non-control worms were lost from a collection of 7,529 experiments. No strain collection lost more than 2 worms. The N2 collection of controls lost 5 worms from its total of 1,218 experiments. The CSV files and MAT statistical-significance files are available for both the primary quality-filtered data sets and the secondary, outlier filtered data sets.

Shapiro-Wilk testing (performed using the 'swtest' function by Ahmed Ben Saida) 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 [51].

Our unfiltered histograms, presented within individual MAT files, were constructed by choosing standard bin resolutions (widths and centres) that resulted in roughly 103 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 centres.

B.6.1 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:

Wt2. The Worm Tracker 2.0 version information.

Video. The video information. The video 'length' is presented as both 'frames' and 'time'. The video 'resolution' is in 'fps' (frames/seconds), pixel 'height' and 'width', the ratio of 'micronsPerPixel', and the codec's 'fourcc' identifier. The video frame 'annotations' are presented for all 'frames' with a 'reference' specifying the annotation's numerical 'id', the 'function' it originated from, and a

'message' describing the meaning of the annotation.

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'.

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.

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:

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'.

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.

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.

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 travelled 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' travelled during this event instance (when available), the 'interTime' till the next event, and the 'interDistance' travelled 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' (centred at the associated feature's values). All histograms also contain the 'resolution' (width) of their bins, whether or not there 'isZeroBin', and whether or not the feature 'isSigned'. 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:

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'.

Sets. The data for the set of experiments is presented as the 'mean', 'stdDev', and 'samples' (the number of experiments) of the collected set. **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.

Table B.1: List of strains. Every assayed strain is listed alongside its genotype and annotations.

No	Genotype	Strain	Outcrossed
1	acc-4(ok2371)III	RB1832	0
2	acd-2(ok1237)I	RB1192	0
3	acd-5(ok2657)I	RB2005	0
4	acr-2(ok1887)X	RB1559	0
5	acr-3(ok2049)X	RB1659	0
6	acr-6(ok3117)I	RB2294	0
7	acr-7(tm863)II	FX863	0
8	acr-9(ok933)X	VC649	0
9	acr-10(ok3064)X	RB2262	0
10	acr-11(ok1345)I	RB1263	0
11	acr-14(ok1155)II	RB1132	0
12	acr-15(ok1214)V	RB1172	0

13	acr-18(ok1285)V	RB1226	0
14	acr-19(ad1674)I	DA1674	1
15	acr-21(ok1314)III	RB1250	0
16	acr-23(ok2804)V	RB2119	0
17	asic-1(ok415)I	RB680	0
18	asic-2(ok289)I	RB557	0
19	bas-1(ad446)III	MT7988	2
20	C11D2.2(ok1565)IV	RB1380	0
21	C24G7.1(ok1822)I	RB1523	0
22	C38D9.2(ok1853)V	RB1543	0
23	cat-2(e1112)II	CB1112	0
24	cat-4(e1141)V	CB1141	0
25	daf-3(e1376)X	CB1376	0
26	daf-5(e1386)II	CB1386	0
27	daf-7(m62)III	DR62	0
28	dat-1(ok157)III	RM2702	6
29	del-1(ok150)X	NC279	4
30	del-4(ok1014)I	RB1064	0
31	del-7(ok1187)IV	RB1156	0
32	del-9(ok2353)X	RB1818	0
33	dnc-1(or404)IV	EU1006	4
34	dop-1(vs100)dop-3(vs106)X	LX705	0
35	dop-1(vs101)X	LX636	4
36	dop-2(vs105)V	LX702	4
37	dop-2(vs105)V; dop-1(vs100)X	LX706	0
38	dop-2(vs105)V; dop-3(vs106)X	LX704	0
39	dop-2(vs105)V; dop-1(vs100)dop-3(vs106)X	LX734	0
40	dop-3(vs106)X	LX703	4
41	dop-4(tm1392)X	FG58	5
42	dpy-20(e1282)IV	CB1282	0
43	eat-16(sa609)I	JT609	3
44	egas-2(ok1477)V	VC975	0
45	egas-3(ok1522)V	RB1356	0
46	egg-5(ok1781)I	VC1295	0
47	egl-1(n487)V	MT1082	1

48	egl-2(n693)V	MT1444	1
49	egl-5(n486)III	MT1081	1
50	egl-6(n592)X	MT1222	0
51	egl-7(n575)III	MT1205	0
52	egl-8(n488)V	MT1083	1
53	egl-9(n586)V	MT1216	0
54	egl-10(md176)V	MT8504	6
55	egl-11(n587)V	MT1217	0
56	egl-12(n602)V	MT1232	0
57	egl-13(n483)X	MT1078	1
58	egl-14(n549)X	MT1179	0
59	egl-15(n484)X	MT1079	1
60	egl-17(e1313)X	CB1313	1
61	egl-18(ok290)IV	JR2370	4
62	egl-19(n2368)IV	MT6129	0
63	egl-20(mu39)IV	CF263	5
64	egl-21(n476)IV	KP2018	5
65	egl-21(n611)IV	MT1241	0
66	egl-23(n601)IV	MT1231	1
67	egl-24(n572)III	MT1202	0
68	egl-27(ok151)II	KS99	6
69	egl-28(n570)II	MT1200	1
70	egl-30(ep271)I	CE1047	2
71	egl-30(n686)goa-1(n1134)I	AQ916	0
72	egl-31(n472)I	MT1067	1
73	egl-32(n155)I	MT155	1
74	egl-33(n151)I	MT151	1
75	egl-36(n728)X	MT1540	1
76	egl-37(n742)II	MT1543	0
77	egl-40(n606)IV	MT1236	0
78	egl-42(n995)II	MT2068	0
79	egl-44(n1080)II	MT2247	0
80	egl-46(n1127)V	MT2316	0
81	egl-47(n1081)V	MT2248	0
82	egl-49(n1107)X	MT2293	0

83	egl-50(n1086)II	AQ2316	5
84	ets-10(gk596)X	VC1340	0
85	F23B2.3(ok1226)IV	RB1177	0
86	F54E4.4(ok2336)X	RB1802	0
87	flp-1(yn2)IV	NY7	0
88	flp-3(ok3265)X	VC2497	0
89	flp-6(ok3056)V	VC2324	0
90	flp-7(ok2625)X	RB1990	0
91	flp-9(ok2730)IV	RB2067	0
92	flp-10(ok2624)IV	RB1989	0
93	flp-11(tm2706)X	FX2706	0
94	flp-12(ok2409)X	RB1863	0
95	flp-13(tm2427)IV	FX2427	0
96	flp-16(ok3085)II	RB2275	0
97	flp-17(ok3587)IV	RB2575	0
98	flp-18(db99)X	AX1410	6
99	flp-19(ok2460)X	RB1902	0
100	flp-20(ok2964)X	RB2188	0
101	flp-21(ok889)V	RB982	0
102	flp-25(gk1016)III	VC1982	0
103	flp-28(gk1075)X	VC2502	0
104	flp-33(gk1038)I	VC2423	0
105	flr-1(ut11)X	JC55	0
106	gar-2(ok520)III	RB756	0
107	gld-1(op236)I	TG34	5
108	gly-2(gk204)I	VC335	0
109	goa-1(sa734)I	DG1856	6
110	gon-2(q362)I	EJ26	4
111	gpa-1(pk15)V	NL332	6
112	gpa-2(pk16)V	NL334	7
113	gpa-3(pk35)V	NL335	7
114	gpa-4(pk381)IV	NL790	6
115	gpa-5(pk376)X	NL1137	6
116	gpa-6(pk480)X	NL1146	6
117	gpa-7(pk610)IV	NL795	6

118	gpa-8(pk345)V	NL1142	6
119	gpa-9(pk438)V	NL793	6
120	gpa-10(pk362)V	NL1147	6
121	gpa-11(pk349)II	NL787	6
122	gpa-12(pk322)X	NL594	6
123	gpa-13(pk1270)V	NL2330	6
124	gpa-14(pk347)I	AQ495	1
125	gpa-15(pk477)I	NL797	6
126	gpa-16(ok2349)I	RB1816	0
127	gpa-17(ok2334)III	RB1800	0
128	gpb-2(sa603)I	JT603	2
129	gpc-1(pk298)X	NL792	6
130	hcf-1(ok559)IV	RB777	0
131	ins-3(ok2488)II	RB1915	0
132	ins-4(ok3534)II	RB2544	0
133	ins-11(tm1053)II	FX1053	5
134	ins-15(ok3444)II	RB2489	0
135	ins-16(ok2919)III	RB2159	0
136	ins-18(ok1672)I	VC1218	1
137	ins-22(ok3616)III	RB2594	0
138	ins-25(ok2773)I	RB2098	0
139	ins-27(ok2474)I	RB1911	0
140	ins-28(ok2722)I	RB2059	0
141	ins-30(ok2343)I	RB1809	0
142	ins-31(ok3543)II	RB2552	0
143	ins-35(ok3297)V	RB2412	0
144	jnk-1(gk7)IV	VC8	0
145	lev-1(x427)IV	ZZ427	4
146	lev-8(x15)X	ZZ15	0
147	lig-4(ok716)III	RB873	0
148	lin-39(n709)III	MT1514	1
149	lon-2(e678)X	CB678	0
150	lov-1(ok522)II	RB753	0
151	mec-4(u253)X; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ908	0
152	mec-7(u448)X; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ1033	0

153	mec-10(e1515)X	CB1515	0
154	mec-10(tm1552)X; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ1413	0
155	mec-10(u20)X	AQ2533	0
156	mec-12(e1605)III; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ1031	0
157	mec-12(u76)III; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ1037	0
158	mec-14(u55)III; bzIs18[pmec-4::YC2.12; lin-15(+)]	AQ1038	0
159	mec-18(u228)X; bzIs17[pmec-4::YC2.12; lin-15(+)]	AQ2649	0
160	mir-124(n4255)IV	MT13292	2
161	mod-1(ok103)V	MT9668	6
162	mod-5(n822)I	MT8944	2
163	nca-2(gk5)III	VC9	0
164	nhr-95(gk836)V	VC1759	0
165	nlp-1(ok1469)X	RB1340	0
166	nlp-2(tm1908)X	FX1908	0
167	nlp-3(ok2688)X	RB2030	0
168	nlp-8(ok1799)I	VC1309	0
169	nlp-12(ok335)IV	RB607	0
170	nlp-14(tm1880)X	FX1880	0
171	nlp-15(ok1512)I	VC1063	0
172	nlp-17(ok3461)IV	RB2498	0
173	nlp-18(ok1557)II	RB1372	0
174	nlp-20(ok1591)IV	RB1396	0
175	npr-1(ad609)X	DA609	2
176	npr-2(ok419)IV	AQ2056	0
177	npr-3(tm1583)IV	FX1583	0
178	npr-4(tm1782)X	AX1743	6
179	npr-5(ok1583)V	AX1745	6
180	npr-7(ok527)X	RB761	0
181	npr-8(tm1553)X	FX1553	0
182	npr-9(tm1652)X	IC683	2
183	npr-10(tm1568)X	FX1568	0
184	npr-11(ok594)X	RB799	0
185	npr-12(tm1498)IV	FX1498	0
186	npr-13(tm1504)V	AQ2153	6
187	npr-20(ok2575)II	RB1958	0

188	ocr-3(ok1559)X	RB1374	0
189	ocr-4(tm2173)IV	FX2173	0
190	ocr-4(vs137)IV	LX950	4
191	ocr-4(vs137)ocr-2(ak47)IV	LX981	0
192	ocr-4(vs137)ocr-2(ak47)IV; ocr-1(ok132)V	LX982	0
193	octr-1(ok371)X	VC224	0
194	odr-3(n2150)V	CX2205	3
195	osm-9(ky10)IV	CX10	0
196	osm-9(ky10)trpa-1(ok999)IV	AQ1422	0
197	pdl-1(gk157)II	VC282	0
198	pkc-1(nj3)V	IK130	6
199	pkg-1(n478)IV	MT1073	1
200	pmk-1(km25)IV	KU25	6
201	pqn-66(ok1507)II	RB1350	0
202	rab-3(y250)II	NM210	2
203	ric-19(ok833)	RB946	0
204	sem-4(ga82)I	EW35	3
205	ser-1(ok345)X	DA1814	10
206	ser-2(pk1357)X	OH313	4
207	ser-4(ok512)III	AQ866	5
208	ser-5(tm2654)I	AQ2197	6
209	ser-6(tm2146)IV	FX2146	0
210	ser-7(tm1325)X	DA2100	10
211	sma-2(e502)III	CB502	0
212	sma-3(e491)III	CB491	0
213	snf-1(ok790)I	RB919	0
214	snf-2(ok147)I	RB641	0
215	snf-4(ok496)II	RB738	0
216	snf-5(ok447)II	RB687	0
217	snf-6(eg28)III	BZ28	4
218	snf-7(ok482)III	RB709	0
219	snf-8(ok349)IV	RB648	0
220	snf-9(ok957)IV	RB1030	0
221	snf-10(hc194)V	BA1093	6
222	snf-11(ok156)V	RM2710	6

223	sng-1(ok234)X	RB503	0
224	spe-41(sy693)III; him-5(e1490)V	PS4330	8
225	srp-8(ok291)V	RB559	0
226	sup-9(n180)II	MT180	0
227	syg-1(ok3640)X	RB2615	0
228	syg-2(ky671)X	CX6391	0
229	T14B1.1(ok1702)X	VC1243	1
230	tbh-1(n3247)X	MT9455	8
231	tdc-1(n3419)II	MT13113	11
232	tom-1(ok285)I	VC223	0
233	tph-1(mg280)II	MT15434	4
234	trp-1(sy690)III	TQ225	7
235	trp-2(gk298)III	VC602	0
236	trp-2(sy691)III	TQ194	8
237	trp-4(sy695)I	TQ296	8
238	trpa-1(ok999)IV	RB1052	0
239	trpa-2(ok3189)I	RB2351	0
240	trpa-2(tm3085)I	FX3085	0
241	trpa-2(tm3092)I	FX3092	0
242	trpl-2(ok2433)IV	RB1883	0
243	tyra-2(tm1846)X	FX1846	0
244	tyra-3(ok325)X	VC125	0
245	unc-1(e1598)X	CB1598	0
246	unc-1(e94)X	CB94	0
247	unc-2(gk366)X	VC854	0
248	unc-2(ox106)X	EG106	0
249	unc-3(e151)X	CB151	0
250	unc-4(e120)II	CB120	0
251	unc-4(gk705)II	VC1528	1
252	unc-7(e5)X	CB5	0
253	unc-8(e15)IV	CB15	0
254	unc-8(e15lb145)IV	MP145	3
255	unc-8(n491n1192)IV	MT2611	0
256	unc-8(e15lb145)IV; del-1(ok150)X	AQ2937	0
257	unc-9(e101)X	CB101	0

258	unc-10(e102)X	CB102	0
259	unc-10(md1117)X	NM1657	5
260	unc-14(e57)I	CB57	0
261	unc-16(e109)III	CB109	0
262	unc-18(e81)X	CB81	0
263	unc-26(m2)IV	DR2	0
264	unc-29(e193)I	CB193	4
265	unc-30(e191)IV	CB845	0
266	unc-31(e169)IV	CB169	0
267	unc-32(e189)III	CB189	0
268	unc-34(e566)V	CB566	1
269	unc-37(e262)I	CB262	0
270	unc-38(e264)I	CB904	4
271	unc-40(n324)I	MT324	0
272	unc-42(e270)V	CB270	0
273	unc-44(e1197)IV	CB1197	0
274	unc-55(e402)I	CB402	0
275	unc-60(e723)V	CB723	0
276	unc-63(ok1075)I	VC731	1
277	unc-69(e587)III	CB587	0
278	unc-75(e950)I	CB950	0
279	unc-76(e911)V	DR96	0
280	unc-77(e625)IV	DR1089	0
281	unc-77(gk9)IV	VC12	0
282	unc-79(e1068)III	CB1068	0
283	unc-80(e1069)V	CB1069	0
284	unc-86(e1416)III	CB1416	0
285	unc-89(e1460)I	CB1460	0
286	unc-89(st85)I	RW85	0
287	unc-98(su130)X	HE130	0
288	unc-101(m1)I	DR1	0
289	unc-103(e1597)III	CB1597	0
290	unc-104(e1265)II	CB1265	0
291	unc-105(ok1432)II	RB1316	0
292	unc-108(n501)I	MT1093	0

293	unc-108(n777)I	MT1656	0
294	unc-115(mn481)X	SP1789	0
295	unc-116(e2310)III	FF41	15
296	unc-118(e2331)X	CB4371	0
297	unc-122(e2520)I	CB4870	2
298	unc-127(hs13)V	HH27	2
299	vab-7(e1562)III	CB1562	0
300	zyg-9(b244)II	DH244	2
301	C. elegans Wild Isolate CB4856 (Hawaii, USA)	CB4856	
302	C. elegans Wild Isolate RC301 (Freiburg, Germany)	RC301	
303	Axenic Liquid Culture LSJ1 (Bristol, UK)	LSJ1	
304	CGC N2 (Bristol, UK)	AQ2947	

Table B.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. Depending on the feature $\pm=\frac{D}{V}$ could mean (+/- = D/V Inside) or (+/- = Toward D/V), $\pm=\frac{F}{B}$ means (+/- = Forward/Backward).

No	Name	Sub div 1	Sub div 2	Units
1	Length			Microns
2	Length	Forward		Microns
3	Length	Paused		Microns
4	Length	Backward		Microns
5	Head Width			Microns
6	Head Width	Forward		Microns
7	Head Width	Paused		Microns
8	Head Width	Backward		Microns
9	Midbody Width			Microns
10	Midbody Width	Forward		Microns
11	Midbody Width	Paused		Microns
12	Midbody Width	Backward		Microns
13	Tail Width			Microns
14	Tail Width	Forward		Microns
15	Tail Width	Paused		Microns
16	Tail Width	Backward		Microns

17	Area			Microns
18	Area	Forward		Microns
19	Area	Paused		Microns
20	Area	Backward		Microns
21	Area/Length			Microns
22	Area/Length	Forward		Microns
23	Area/Length	Paused		Microns
24	Area/Length	Backward		Microns
25	Width/Length			No Units
26	Width/Length	Forward		No Units
27	Width/Length	Paused		No Units
28	Width/Length	Backward		No Units
29	Head Bend Mean ($\pm$=$\frac{D}{V}$)			Degrees
30	Head Bend Mean ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
31	Head Bend Mean ($\pm$=$\frac{D}{V}$)		Positive	Degrees
32	Head Bend Mean ($\pm$=$\frac{D}{V}$)		Negative	Degrees
33	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Forward		Degrees
34	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
35	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees
36	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Negative	Degrees
37	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Paused		Degrees
38	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Degrees
39	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
40	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
41	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Backward		Degrees
42	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Degrees
43	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Positive	Degrees
44	Head Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Negative	Degrees
45	Neck Bend Mean ($\pm$=$\frac{D}{V}$)			Degrees
46	Neck Bend Mean ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
47	Neck Bend Mean ($\pm$=$\frac{D}{V}$)		Positive	Degrees
48	Neck Bend Mean ($\pm$=$\frac{D}{V}$)		Negative	Degrees
49	Neck Bend Mean ($\pm$=$\frac{D}{V}$)	Forward		Degrees
50	Neck Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
51	Neck Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees

52	Neck Bend Mean ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
53	Neck Bend Mean ($\pm=\frac{D}{V}$)	Paused		Degrees
54	Neck Bend Mean ($\pm=\frac{D}{V}$)	Paused	Absolute	Degrees
55	Neck Bend Mean ($\pm=\frac{D}{V}$)	Paused	Positive	Degrees
56	Neck Bend Mean ($\pm=\frac{D}{V}$)	Paused	Negative	Degrees
57	Neck Bend Mean ($\pm=\frac{D}{V}$)	Backward		Degrees
58	Neck Bend Mean ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
59	Neck Bend Mean ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
60	Neck Bend Mean ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
61	Midbody Bend Mean ($\pm=\frac{D}{V}$)			Degrees
62	Midbody Bend Mean ($\pm=\frac{D}{V}$)		Absolute	Degrees
63	Midbody Bend Mean ($\pm=\frac{D}{V}$)		Positive	Degrees
64	Midbody Bend Mean ($\pm=\frac{D}{V}$)		Negative	Degrees
65	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Forward		Degrees
66	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
67	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
68	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
69	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Paused		Degrees
70	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Paused	Absolute	Degrees
71	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Paused	Positive	Degrees
72	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Paused	Negative	Degrees
73	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Backward		Degrees
74	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
75	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
76	Midbody Bend Mean ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
77	Hips Bend Mean ($\pm=\frac{D}{V}$)			Degrees
78	Hips Bend Mean ($\pm=\frac{D}{V}$)		Absolute	Degrees
79	Hips Bend Mean ($\pm=\frac{D}{V}$)		Positive	Degrees
80	Hips Bend Mean ($\pm=\frac{D}{V}$)		Negative	Degrees
81	Hips Bend Mean ($\pm=\frac{D}{V}$)	Forward		Degrees
82	Hips Bend Mean ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
83	Hips Bend Mean ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
84	Hips Bend Mean ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
85	Hips Bend Mean ($\pm=\frac{D}{V}$)	Paused		Degrees
86	Hips Bend Mean ($\pm=\frac{D}{V}$)	Paused	Absolute	Degrees

87	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
88	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
89	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Backward		Degrees
90	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Degrees
91	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Positive	Degrees
92	Hips Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Negative	Degrees
93	Tail Bend Mean ($\pm$=$\frac{D}{V}$)			Degrees
94	Tail Bend Mean ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
95	Tail Bend Mean ($\pm$=$\frac{D}{V}$)		Positive	Degrees
96	Tail Bend Mean ($\pm$=$\frac{D}{V}$)		Negative	Degrees
97	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Forward		Degrees
98	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
99	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees
100	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Forward	Negative	Degrees
101	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Paused		Degrees
102	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Degrees
103	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
104	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
105	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Backward		Degrees
106	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Degrees
107	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Positive	Degrees
108	Tail Bend Mean ($\pm$=$\frac{D}{V}$)	Backward	Negative	Degrees
109	Head Bend S.D. ($\pm$=$\frac{D}{V}$)			Degrees
110	Head Bend S.D. ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
111	Head Bend S.D. ($\pm$=$\frac{D}{V}$)		Positive	Degrees
112	Head Bend S.D. ($\pm$=$\frac{D}{V}$)		Negative	Degrees
113	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward		Degrees
114	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
115	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees
116	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Degrees
117	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused		Degrees
118	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Degrees
119	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
120	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
121	Head Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward		Degrees

122	Head Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
123	Head Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
124	Head Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
125	Neck Bend S.D. ($\pm=\frac{D}{V}$)			Degrees
126	Neck Bend S.D. ($\pm=\frac{D}{V}$)		Absolute	Degrees
127	Neck Bend S.D. ($\pm=\frac{D}{V}$)		Positive	Degrees
128	Neck Bend S.D. ($\pm=\frac{D}{V}$)		Negative	Degrees
129	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Forward		Degrees
130	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
131	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
132	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
133	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Paused		Degrees
134	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Absolute	Degrees
135	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Positive	Degrees
136	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Negative	Degrees
137	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Backward		Degrees
138	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
139	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
140	Neck Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
141	Midbody Bend S.D. ($\pm=\frac{D}{V}$)			Degrees
142	Midbody Bend S.D. ($\pm=\frac{D}{V}$)		Absolute	Degrees
143	Midbody Bend S.D. ($\pm=\frac{D}{V}$)		Positive	Degrees
144	Midbody Bend S.D. ($\pm=\frac{D}{V}$)		Negative	Degrees
145	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Forward		Degrees
146	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
147	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
148	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
149	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Paused		Degrees
150	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Absolute	Degrees
151	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Positive	Degrees
152	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Paused	Negative	Degrees
153	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Backward		Degrees
154	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
155	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
156	Midbody Bend S.D. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees

157	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)			Degrees
158	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
159	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)		Positive	Degrees
160	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)		Negative	Degrees
161	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward		Degrees
162	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
163	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees
164	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Degrees
165	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused		Degrees
166	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Degrees
167	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
168	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
169	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward		Degrees
170	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Degrees
171	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Degrees
172	Hips Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Degrees
173	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)			Degrees
174	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)		Absolute	Degrees
175	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)		Positive	Degrees
176	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)		Negative	Degrees
177	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward		Degrees
178	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Degrees
179	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Degrees
180	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Degrees
181	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused		Degrees
182	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Degrees
183	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Degrees
184	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Degrees
185	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward		Degrees
186	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Degrees
187	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Degrees
188	Tail Bend S.D. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Degrees
189	Max Amptd.			Microns
190	Max Amptd.	Forward		Microns
191	Max Amptd.	Paused		Microns

192	Max Amptd.	Backward		Microns
193	Amptd. Ratio			No Units
194	Amptd. Ratio	Forward		No Units
195	Amptd. Ratio	Paused		No Units
196	Amptd. Ratio	Backward		No Units
197	Primary Wavelength			Microns
198	Primary Wavelength	Forward		Microns
199	Primary Wavelength	Paused		Microns
200	Primary Wavelength	Backward		Microns
201	Secondary Wavelength			Microns
202	Secondary Wavelength	Forward		Microns
203	Secondary Wavelength	Paused		Microns
204	Secondary Wavelength	Backward		Microns
205	Track Length			Microns
206	Track Length	Forward		Microns
207	Track Length	Paused		Microns
208	Track Length	Backward		Microns
209	Eccentricity			No Units
210	Eccentricity	Forward		No Units
211	Eccentricity	Paused		No Units
212	Eccentricity	Backward		No Units
213	Bend Count			Counts
214	Bend Count	Forward		Counts
215	Bend Count	Paused		Counts
216	Bend Count	Backward		Counts
217	Coiling Events	Freq.		Hz
218	Coiling Events	Time Ratio		No Units
219	Coiling Events	Time		Seconds
220	Coiling Events	Inter Time		Seconds
221	Coiling Events	Inter Distance		Microns
222	Tail-To-Head Orientation			Degrees
223	Tail-To-Head Orientation		Absolute	Degrees
224	Tail-To-Head Orientation		Positive	Degrees
225	Tail-To-Head Orientation		Negative	Degrees
226	Tail-To-Head Orientation	Forward		Degrees

227	Tail-To-Head Orientation	Forward	Absolute	Degrees
228	Tail-To-Head Orientation	Forward	Positive	Degrees
229	Tail-To-Head Orientation	Forward	Negative	Degrees
230	Tail-To-Head Orientation	Paused		Degrees
231	Tail-To-Head Orientation	Paused	Absolute	Degrees
232	Tail-To-Head Orientation	Paused	Positive	Degrees
233	Tail-To-Head Orientation	Paused	Negative	Degrees
234	Tail-To-Head Orientation	Backward		Degrees
235	Tail-To-Head Orientation	Backward	Absolute	Degrees
236	Tail-To-Head Orientation	Backward	Positive	Degrees
237	Tail-To-Head Orientation	Backward	Negative	Degrees
238	Head Orientation			Degrees
239	Head Orientation		Absolute	Degrees
240	Head Orientation		Positive	Degrees
241	Head Orientation		Negative	Degrees
242	Head Orientation	Forward		Degrees
243	Head Orientation	Forward	Absolute	Degrees
244	Head Orientation	Forward	Positive	Degrees
245	Head Orientation	Forward	Negative	Degrees
246	Head Orientation	Paused		Degrees
247	Head Orientation	Paused	Absolute	Degrees
248	Head Orientation	Paused	Positive	Degrees
249	Head Orientation	Paused	Negative	Degrees
250	Head Orientation	Backward		Degrees
251	Head Orientation	Backward	Absolute	Degrees
252	Head Orientation	Backward	Positive	Degrees
253	Head Orientation	Backward	Negative	Degrees
254	Tail Orientation			Degrees
255	Tail Orientation		Absolute	Degrees
256	Tail Orientation		Positive	Degrees
257	Tail Orientation		Negative	Degrees
258	Tail Orientation	Forward		Degrees
259	Tail Orientation	Forward	Absolute	Degrees
260	Tail Orientation	Forward	Positive	Degrees
261	Tail Orientation	Forward	Negative	Degrees

262	Tail Orientation	Paused		Degrees
263	Tail Orientation	Paused	Absolute	Degrees
264	Tail Orientation	Paused	Positive	Degrees
265	Tail Orientation	Paused	Negative	Degrees
266	Tail Orientation	Backward		Degrees
267	Tail Orientation	Backward	Absolute	Degrees
268	Tail Orientation	Backward	Positive	Degrees
269	Tail Orientation	Backward	Negative	Degrees
270	Eigen Projection 1			No Units
271	Eigen Projection 1		Absolute	No Units
272	Eigen Projection 1		Positive	No Units
273	Eigen Projection 1		Negative	No Units
274	Eigen Projection 1	Forward		No Units
275	Eigen Projection 1	Forward	Absolute	No Units
276	Eigen Projection 1	Forward	Positive	No Units
277	Eigen Projection 1	Forward	Negative	No Units
278	Eigen Projection 1	Paused		No Units
279	Eigen Projection 1	Paused	Absolute	No Units
280	Eigen Projection 1	Paused	Positive	No Units
281	Eigen Projection 1	Paused	Negative	No Units
282	Eigen Projection 1	Backward		No Units
283	Eigen Projection 1	Backward	Absolute	No Units
284	Eigen Projection 1	Backward	Positive	No Units
285	Eigen Projection 1	Backward	Negative	No Units
286	Eigen Projection 2			No Units
287	Eigen Projection 2		Absolute	No Units
288	Eigen Projection 2		Positive	No Units
289	Eigen Projection 2		Negative	No Units
290	Eigen Projection 2	Forward		No Units
291	Eigen Projection 2	Forward	Absolute	No Units
292	Eigen Projection 2	Forward	Positive	No Units
293	Eigen Projection 2	Forward	Negative	No Units
294	Eigen Projection 2	Paused		No Units
295	Eigen Projection 2	Paused	Absolute	No Units
296	Eigen Projection 2	Paused	Positive	No Units

297	Eigen Projection 2	Paused	Negative	No Units
298	Eigen Projection 2	Backward		No Units
299	Eigen Projection 2	Backward	Absolute	No Units
300	Eigen Projection 2	Backward	Positive	No Units
301	Eigen Projection 2	Backward	Negative	No Units
302	Eigen Projection 3			No Units
303	Eigen Projection 3		Absolute	No Units
304	Eigen Projection 3		Positive	No Units
305	Eigen Projection 3		Negative	No Units
306	Eigen Projection 3	Forward		No Units
307	Eigen Projection 3	Forward	Absolute	No Units
308	Eigen Projection 3	Forward	Positive	No Units
309	Eigen Projection 3	Forward	Negative	No Units
310	Eigen Projection 3	Paused		No Units
311	Eigen Projection 3	Paused	Absolute	No Units
312	Eigen Projection 3	Paused	Positive	No Units
313	Eigen Projection 3	Paused	Negative	No Units
314	Eigen Projection 3	Backward		No Units
315	Eigen Projection 3	Backward	Absolute	No Units
316	Eigen Projection 3	Backward	Positive	No Units
317	Eigen Projection 3	Backward	Negative	No Units
318	Eigen Projection 4			No Units
319	Eigen Projection 4		Absolute	No Units
320	Eigen Projection 4		Positive	No Units
321	Eigen Projection 4		Negative	No Units
322	Eigen Projection 4	Forward		No Units
323	Eigen Projection 4	Forward	Absolute	No Units
324	Eigen Projection 4	Forward	Positive	No Units
325	Eigen Projection 4	Forward	Negative	No Units
326	Eigen Projection 4	Paused		No Units
327	Eigen Projection 4	Paused	Absolute	No Units
328	Eigen Projection 4	Paused	Positive	No Units
329	Eigen Projection 4	Paused	Negative	No Units
330	Eigen Projection 4	Backward		No Units
331	Eigen Projection 4	Backward	Absolute	No Units

332	Eigen Projection 4	Backward	Positive	No Units
333	Eigen Projection 4	Backward	Negative	No Units
334	Eigen Projection 5			No Units
335	Eigen Projection 5		Absolute	No Units
336	Eigen Projection 5		Positive	No Units
337	Eigen Projection 5		Negative	No Units
338	Eigen Projection 5	Forward		No Units
339	Eigen Projection 5	Forward	Absolute	No Units
340	Eigen Projection 5	Forward	Positive	No Units
341	Eigen Projection 5	Forward	Negative	No Units
342	Eigen Projection 5	Paused		No Units
343	Eigen Projection 5	Paused	Absolute	No Units
344	Eigen Projection 5	Paused	Positive	No Units
345	Eigen Projection 5	Paused	Negative	No Units
346	Eigen Projection 5	Backward		No Units
347	Eigen Projection 5	Backward	Absolute	No Units
348	Eigen Projection 5	Backward	Positive	No Units
349	Eigen Projection 5	Backward	Negative	No Units
350	Eigen Projection 6			No Units
351	Eigen Projection 6		Absolute	No Units
352	Eigen Projection 6		Positive	No Units
353	Eigen Projection 6		Negative	No Units
354	Eigen Projection 6	Forward		No Units
355	Eigen Projection 6	Forward	Absolute	No Units
356	Eigen Projection 6	Forward	Positive	No Units
357	Eigen Projection 6	Forward	Negative	No Units
358	Eigen Projection 6	Paused		No Units
359	Eigen Projection 6	Paused	Absolute	No Units
360	Eigen Projection 6	Paused	Positive	No Units
361	Eigen Projection 6	Paused	Negative	No Units
362	Eigen Projection 6	Backward		No Units
363	Eigen Projection 6	Backward	Absolute	No Units
364	Eigen Projection 6	Backward	Positive	No Units
365	Eigen Projection 6	Backward	Negative	No Units
366	Forward Motion	Freq.		Hz

367	Forward Motion	Time Ratio		No Units
368	Forward Motion	Distance Ratio		No Units
369	Forward Motion	Time		Seconds
370	Forward Motion	Distance		Microns
371	Forward Motion	Inter Time		Seconds
372	Forward Motion	Inter Distance		Microns
373	Paused Motion	Freq.		Hz
374	Paused Motion	Time Ratio		No Units
375	Paused Motion	Distance Ratio		No Units
376	Paused Motion	Time		Seconds
377	Paused Motion	Distance		Microns
378	Paused Motion	Inter Time		Seconds
379	Paused Motion	Inter Distance		Microns
380	Backward Motion	Freq.		Hz
381	Backward Motion	Time Ratio		No Units
382	Backward Motion	Distance Ratio		No Units
383	Backward Motion	Time		Seconds
384	Backward Motion	Distance		Microns
385	Backward Motion	Inter Time		Seconds
386	Backward Motion	Inter Distance		Microns
387	Head Tip Speed ($\pm=\frac{F}{B}$)			μ/Sec
388	Head Tip Speed ($\pm=\frac{F}{B}$)		Absolute	μ/Sec
389	Head Tip Speed ($\pm=\frac{F}{B}$)		Positive	μ/Sec
390	Head Tip Speed ($\pm=\frac{F}{B}$)		Negative	μ/Sec
391	Head Tip Speed ($\pm=\frac{F}{B}$)	Forward		μ/Sec
392	Head Tip Speed ($\pm=\frac{F}{B}$)	Forward	Absolute	μ/Sec
393	Head Tip Speed ($\pm=\frac{F}{B}$)	Forward	Positive	μ/Sec
394	Head Tip Speed ($\pm=\frac{F}{B}$)	Forward	Negative	μ/Sec
395	Head Tip Speed ($\pm=\frac{F}{B}$)	Paused		μ/Sec
396	Head Tip Speed ($\pm=\frac{F}{B}$)	Paused	Absolute	μ/Sec
397	Head Tip Speed ($\pm=\frac{F}{B}$)	Paused	Positive	μ/Sec
398	Head Tip Speed ($\pm=\frac{F}{B}$)	Paused	Negative	μ/Sec
399	Head Tip Speed ($\pm=\frac{F}{B}$)	Backward		μ/Sec
400	Head Tip Speed ($\pm=\frac{F}{B}$)	Backward	Absolute	μ/Sec
401	Head Tip Speed ($\pm=\frac{F}{B}$)	Backward	Positive	μ/Sec

402	Head Tip Speed ($\pm=\frac{F}{B}$)	Backward	Negative	μ/Sec
403	Head Speed ($\pm=\frac{F}{B}$)			μ/Sec
404	Head Speed ($\pm=\frac{F}{B}$)		Absolute	μ/Sec
405	Head Speed ($\pm=\frac{F}{B}$)		Positive	μ/Sec
406	Head Speed ($\pm=\frac{F}{B}$)		Negative	μ/Sec
407	Head Speed ($\pm=\frac{F}{B}$)	Forward		μ/Sec
408	Head Speed ($\pm=\frac{F}{B}$)	Forward	Absolute	μ/Sec
409	Head Speed ($\pm=\frac{F}{B}$)	Forward	Positive	μ/Sec
410	Head Speed ($\pm=\frac{F}{B}$)	Forward	Negative	μ/Sec
411	Head Speed ($\pm=\frac{F}{B}$)	Paused		μ/Sec
412	Head Speed ($\pm=\frac{F}{B}$)	Paused	Absolute	μ/Sec
413	Head Speed ($\pm=\frac{F}{B}$)	Paused	Positive	μ/Sec
414	Head Speed ($\pm=\frac{F}{B}$)	Paused	Negative	μ/Sec
415	Head Speed ($\pm=\frac{F}{B}$)	Backward		μ/Sec
416	Head Speed ($\pm=\frac{F}{B}$)	Backward	Absolute	μ/Sec
417	Head Speed ($\pm=\frac{F}{B}$)	Backward	Positive	μ/Sec
418	Head Speed ($\pm=\frac{F}{B}$)	Backward	Negative	μ/Sec
419	Midbody Speed ($\pm=\frac{F}{B}$)			μ/Sec
420	Midbody Speed ($\pm=\frac{F}{B}$)		Absolute	μ/Sec
421	Midbody Speed ($\pm=\frac{F}{B}$)		Positive	μ/Sec
422	Midbody Speed ($\pm=\frac{F}{B}$)		Negative	μ/Sec
423	Midbody Speed ($\pm=\frac{F}{B}$)	Forward		μ/Sec
424	Midbody Speed ($\pm=\frac{F}{B}$)	Forward	Absolute	μ/Sec
425	Midbody Speed ($\pm=\frac{F}{B}$)	Forward	Positive	μ/Sec
426	Midbody Speed ($\pm=\frac{F}{B}$)	Forward	Negative	μ/Sec
427	Midbody Speed ($\pm=\frac{F}{B}$)	Paused		μ/Sec
428	Midbody Speed ($\pm=\frac{F}{B}$)	Paused	Absolute	μ/Sec
429	Midbody Speed ($\pm=\frac{F}{B}$)	Paused	Positive	μ/Sec
430	Midbody Speed ($\pm=\frac{F}{B}$)	Paused	Negative	μ/Sec
431	Midbody Speed ($\pm=\frac{F}{B}$)	Backward		μ/Sec
432	Midbody Speed ($\pm=\frac{F}{B}$)	Backward	Absolute	μ/Sec
433	Midbody Speed ($\pm=\frac{F}{B}$)	Backward	Positive	μ/Sec
434	Midbody Speed ($\pm=\frac{F}{B}$)	Backward	Negative	μ/Sec
435	Tail Speed ($\pm=\frac{F}{B}$)			μ/Sec
436	Tail Speed ($\pm=\frac{F}{B}$)		Absolute	μ/Sec

437	Tail Speed ($\pm=\frac{F}{B}$)		Positive	μ/Sec
438	Tail Speed ($\pm=\frac{F}{B}$)		Negative	μ/Sec
439	Tail Speed ($\pm=\frac{F}{B}$)	Forward		μ/Sec
440	Tail Speed ($\pm=\frac{F}{B}$)	Forward	Absolute	μ/Sec
441	Tail Speed ($\pm=\frac{F}{B}$)	Forward	Positive	μ/Sec
442	Tail Speed ($\pm=\frac{F}{B}$)	Forward	Negative	μ/Sec
443	Tail Speed ($\pm=\frac{F}{B}$)	Paused		μ/Sec
444	Tail Speed ($\pm=\frac{F}{B}$)	Paused	Absolute	μ/Sec
445	Tail Speed ($\pm=\frac{F}{B}$)	Paused	Positive	μ/Sec
446	Tail Speed ($\pm=\frac{F}{B}$)	Paused	Negative	μ/Sec
447	Tail Speed ($\pm=\frac{F}{B}$)	Backward		μ/Sec
448	Tail Speed ($\pm=\frac{F}{B}$)	Backward	Absolute	μ/Sec
449	Tail Speed ($\pm=\frac{F}{B}$)	Backward	Positive	μ/Sec
450	Tail Speed ($\pm=\frac{F}{B}$)	Backward	Negative	μ/Sec
451	Tail Tip Speed ($\pm=\frac{F}{B}$)			μ/Sec
452	Tail Tip Speed ($\pm=\frac{F}{B}$)		Absolute	μ/Sec
453	Tail Tip Speed ($\pm=\frac{F}{B}$)		Positive	μ/Sec
454	Tail Tip Speed ($\pm=\frac{F}{B}$)		Negative	μ/Sec
455	Tail Tip Speed ($\pm=\frac{F}{B}$)	Forward		μ/Sec
456	Tail Tip Speed ($\pm=\frac{F}{B}$)	Forward	Absolute	μ/Sec
457	Tail Tip Speed ($\pm=\frac{F}{B}$)	Forward	Positive	μ/Sec
458	Tail Tip Speed ($\pm=\frac{F}{B}$)	Forward	Negative	μ/Sec
459	Tail Tip Speed ($\pm=\frac{F}{B}$)	Paused		μ/Sec
460	Tail Tip Speed ($\pm=\frac{F}{B}$)	Paused	Absolute	μ/Sec
461	Tail Tip Speed ($\pm=\frac{F}{B}$)	Paused	Positive	μ/Sec
462	Tail Tip Speed ($\pm=\frac{F}{B}$)	Paused	Negative	μ/Sec
463	Tail Tip Speed ($\pm=\frac{F}{B}$)	Backward		μ/Sec
464	Tail Tip Speed ($\pm=\frac{F}{B}$)	Backward	Absolute	μ/Sec
465	Tail Tip Speed ($\pm=\frac{F}{B}$)	Backward	Positive	μ/Sec
466	Tail Tip Speed ($\pm=\frac{F}{B}$)	Backward	Negative	μ/Sec
467	Head Tip Motion Dir. ($\pm=\frac{D}{V}$)			Deg/Sec
468	Head Tip Motion Dir. ($\pm=\frac{D}{V}$)		Absolute	Deg/Sec
469	Head Tip Motion Dir. ($\pm=\frac{D}{V}$)		Positive	Deg/Sec
470	Head Tip Motion Dir. ($\pm=\frac{D}{V}$)		Negative	Deg/Sec
471	Head Tip Motion Dir. ($\pm=\frac{D}{V}$)	Forward		Deg/Sec

472	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Deg/Sec
473	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Deg/Sec
474	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Deg/Sec
475	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused		Deg/Sec
476	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Deg/Sec
477	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Deg/Sec
478	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Deg/Sec
479	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward		Deg/Sec
480	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Deg/Sec
481	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Deg/Sec
482	Head Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Deg/Sec
483	Head Motion Dir. ($\pm$=$\frac{D}{V}$)			Deg/Sec
484	Head Motion Dir. ($\pm$=$\frac{D}{V}$)		Absolute	Deg/Sec
485	Head Motion Dir. ($\pm$=$\frac{D}{V}$)		Positive	Deg/Sec
486	Head Motion Dir. ($\pm$=$\frac{D}{V}$)		Negative	Deg/Sec
487	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward		Deg/Sec
488	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Deg/Sec
489	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Deg/Sec
490	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Deg/Sec
491	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused		Deg/Sec
492	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Deg/Sec
493	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Deg/Sec
494	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Deg/Sec
495	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward		Deg/Sec
496	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Deg/Sec
497	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Deg/Sec
498	Head Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Deg/Sec
499	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)			Deg/Sec
500	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)		Absolute	Deg/Sec
501	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)		Positive	Deg/Sec
502	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)		Negative	Deg/Sec
503	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward		Deg/Sec
504	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Deg/Sec
505	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Deg/Sec
506	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Deg/Sec

507	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused		Deg/Sec
508	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Deg/Sec
509	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Deg/Sec
510	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Deg/Sec
511	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward		Deg/Sec
512	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Deg/Sec
513	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Deg/Sec
514	Midbody Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Deg/Sec
515	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)			Deg/Sec
516	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)		Absolute	Deg/Sec
517	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)		Positive	Deg/Sec
518	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)		Negative	Deg/Sec
519	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward		Deg/Sec
520	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Deg/Sec
521	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Deg/Sec
522	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Deg/Sec
523	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused		Deg/Sec
524	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Deg/Sec
525	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Deg/Sec
526	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Negative	Deg/Sec
527	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward		Deg/Sec
528	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Absolute	Deg/Sec
529	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Positive	Deg/Sec
530	Tail Motion Dir. ($\pm$=$\frac{D}{V}$)	Backward	Negative	Deg/Sec
531	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)			Deg/Sec
532	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)		Absolute	Deg/Sec
533	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)		Positive	Deg/Sec
534	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)		Negative	Deg/Sec
535	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward		Deg/Sec
536	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Absolute	Deg/Sec
537	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Positive	Deg/Sec
538	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Forward	Negative	Deg/Sec
539	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused		Deg/Sec
540	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Absolute	Deg/Sec
541	Tail Tip Motion Dir. ($\pm$=$\frac{D}{V}$)	Paused	Positive	Deg/Sec

542	Tail Tip Motion Dir. ($\pm=\frac{D}{V}$)	Paused	Negative	Deg/Sec
543	Tail Tip Motion Dir. ($\pm=\frac{D}{V}$)	Backward		Deg/Sec
544	Tail Tip Motion Dir. ($\pm=\frac{D}{V}$)	Backward	Absolute	Deg/Sec
545	Tail Tip Motion Dir. ($\pm=\frac{D}{V}$)	Backward	Positive	Deg/Sec
546	Tail Tip Motion Dir. ($\pm=\frac{D}{V}$)	Backward	Negative	Deg/Sec
547	Foraging Amptd. ($\pm=\frac{D}{V}$)			Microns
548	Foraging Amptd. ($\pm=\frac{D}{V}$)		Absolute	Microns
549	Foraging Amptd. ($\pm=\frac{D}{V}$)		Positive	Microns
550	Foraging Amptd. ($\pm=\frac{D}{V}$)		Negative	Microns
551	Foraging Amptd. ($\pm=\frac{D}{V}$)	Forward		Microns
552	Foraging Amptd. ($\pm=\frac{D}{V}$)	Forward	Absolute	Microns
553	Foraging Amptd. ($\pm=\frac{D}{V}$)	Forward	Positive	Microns
554	Foraging Amptd. ($\pm=\frac{D}{V}$)	Forward	Negative	Microns
555	Foraging Amptd. ($\pm=\frac{D}{V}$)	Paused		Microns
556	Foraging Amptd. ($\pm=\frac{D}{V}$)	Paused	Absolute	Microns
557	Foraging Amptd. ($\pm=\frac{D}{V}$)	Paused	Positive	Microns
558	Foraging Amptd. ($\pm=\frac{D}{V}$)	Paused	Negative	Microns
559	Foraging Amptd. ($\pm=\frac{D}{V}$)	Backward		Microns
560	Foraging Amptd. ($\pm=\frac{D}{V}$)	Backward	Absolute	Microns
561	Foraging Amptd. ($\pm=\frac{D}{V}$)	Backward	Positive	Microns
562	Foraging Amptd. ($\pm=\frac{D}{V}$)	Backward	Negative	Microns
563	Head Crawling Amptd. ($\pm=\frac{D}{V}$)			Degrees
564	Head Crawling Amptd. ($\pm=\frac{D}{V}$)		Absolute	Degrees
565	Head Crawling Amptd. ($\pm=\frac{D}{V}$)		Positive	Degrees
566	Head Crawling Amptd. ($\pm=\frac{D}{V}$)		Negative	Degrees
567	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward		Degrees
568	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
569	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
570	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
571	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward		Degrees
572	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
573	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
574	Head Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
575	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)			Degrees
576	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)		Absolute	Degrees

577	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)		Positive	Degrees
578	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)		Negative	Degrees
579	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward		Degrees
580	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
581	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
582	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
583	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward		Degrees
584	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
585	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
586	Midbody Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
587	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)			Degrees
588	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)		Absolute	Degrees
589	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)		Positive	Degrees
590	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)		Negative	Degrees
591	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward		Degrees
592	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Absolute	Degrees
593	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Positive	Degrees
594	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Forward	Negative	Degrees
595	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward		Degrees
596	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Absolute	Degrees
597	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Positive	Degrees
598	Tail Crawling Amptd. ($\pm=\frac{D}{V}$)	Backward	Negative	Degrees
599	Foraging Speed			Deg/Sec
600	Foraging Speed		Absolute	Deg/Sec
601	Foraging Speed		Positive	Deg/Sec
602	Foraging Speed		Negative	Deg/Sec
603	Foraging Speed	Forward		Deg/Sec
604	Foraging Speed	Forward	Absolute	Deg/Sec
605	Foraging Speed	Forward	Positive	Deg/Sec
606	Foraging Speed	Forward	Negative	Deg/Sec
607	Foraging Speed	Paused		Deg/Sec
608	Foraging Speed	Paused	Absolute	Deg/Sec
609	Foraging Speed	Paused	Positive	Deg/Sec
610	Foraging Speed	Paused	Negative	Deg/Sec
611	Foraging Speed	Backward		Deg/Sec

612	Foraging Speed	Backward	Absolute	Deg/Sec
613	Foraging Speed	Backward	Positive	Deg/Sec
614	Foraging Speed	Backward	Negative	Deg/Sec
615	Head Crawling Freq. ($\pm=\frac{D}{V}$)			Hz
616	Head Crawling Freq. ($\pm=\frac{D}{V}$)		Absolute	Hz
617	Head Crawling Freq. ($\pm=\frac{D}{V}$)		Positive	Hz
618	Head Crawling Freq. ($\pm=\frac{D}{V}$)		Negative	Hz
619	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Forward		Hz
620	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Absolute	Hz
621	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Positive	Hz
622	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Negative	Hz
623	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Backward		Hz
624	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Absolute	Hz
625	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Positive	Hz
626	Head Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Negative	Hz
627	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)			Hz
628	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)		Absolute	Hz
629	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)		Positive	Hz
630	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)		Negative	Hz
631	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Forward		Hz
632	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Absolute	Hz
633	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Positive	Hz
634	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Negative	Hz
635	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Backward		Hz
636	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Absolute	Hz
637	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Positive	Hz
638	Midbody Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Negative	Hz
639	Tail Crawling Freq. ($\pm=\frac{D}{V}$)			Hz
640	Tail Crawling Freq. ($\pm=\frac{D}{V}$)		Absolute	Hz
641	Tail Crawling Freq. ($\pm=\frac{D}{V}$)		Positive	Hz
642	Tail Crawling Freq. ($\pm=\frac{D}{V}$)		Negative	Hz
643	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Forward		Hz
644	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Absolute	Hz
645	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Positive	Hz
646	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Forward	Negative	Hz

647	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Backward		Hz
648	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Absolute	Hz
649	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Positive	Hz
650	Tail Crawling Freq. ($\pm=\frac{D}{V}$)	Backward	Negative	Hz
651	Omega Turn Events	Freq.		Hz
652	Omega Turn Events	Time Ratio		No Units
653	Omega Turn Events	Time		Seconds
654	Omega Turn Events	Time	Absolute	Seconds
655	Omega Turn Events	Time	Positive	Seconds
656	Omega Turn Events	Time	Negative	Seconds
657	Omega Turn Events	Inter Time		Seconds
658	Omega Turn Events	Inter Time	Absolute	Seconds
659	Omega Turn Events	Inter Time	Positive	Seconds
660	Omega Turn Events	Inter Time	Negative	Seconds
661	Omega Turn Events	Inter Distance		Microns
662	Omega Turn Events	Inter Distance	Absolute	Microns
663	Omega Turn Events	Inter Distance	Positive	Microns
664	Omega Turn Events	Inter Distance	Negative	Microns
665	Upsilon Turn Events	Freq.		Hz
666	Upsilon Turn Events	Time Ratio		No Units
667	Upsilon Turn Events	Time		Seconds
668	Upsilon Turn Events	Time	Absolute	Seconds
669	Upsilon Turn Events	Time	Positive	Seconds
670	Upsilon Turn Events	Time	Negative	Seconds
671	Upsilon Turn Events	Inter Time		Seconds
672	Upsilon Turn Events	Inter Time	Absolute	Seconds
673	Upsilon Turn Events	Inter Time	Positive	Seconds
674	Upsilon Turn Events	Inter Time	Negative	Seconds
675	Upsilon Turn Events	Inter Distance		Microns
676	Upsilon Turn Events	Inter Distance	Absolute	Microns
677	Upsilon Turn Events	Inter Distance	Positive	Microns
678	Upsilon Turn Events	Inter Distance	Negative	Microns
679	Path Range			Microns
680	Path Range	Forward		Microns
681	Path Range	Paused		Microns

682	Path Range	Backward		Microns
683	Worm Dwelling			Seconds
684	Head Dwelling			Seconds
685	Midbody Dwelling			Seconds
686	Tail Dwelling			Seconds
687	Path Curvature ($\pm=\frac{D}{V}$)			Rad/μ
688	Path Curvature ($\pm=\frac{D}{V}$)		Absolute	Rad/μ
689	Path Curvature ($\pm=\frac{D}{V}$)		Positive	Rad/μ
690	Path Curvature ($\pm=\frac{D}{V}$)		Negative	Rad/μ
691	Path Curvature ($\pm=\frac{D}{V}$)	Forward		Rad/μ
692	Path Curvature ($\pm=\frac{D}{V}$)	Forward	Absolute	Rad/μ
693	Path Curvature ($\pm=\frac{D}{V}$)	Forward	Positive	Rad/μ
694	Path Curvature ($\pm=\frac{D}{V}$)	Forward	Negative	Rad/μ
695	Path Curvature ($\pm=\frac{D}{V}$)	Paused		Rad/μ
696	Path Curvature ($\pm=\frac{D}{V}$)	Paused	Absolute	Rad/μ
697	Path Curvature ($\pm=\frac{D}{V}$)	Paused	Positive	Rad/μ
698	Path Curvature ($\pm=\frac{D}{V}$)	Paused	Negative	Rad/μ
699	Path Curvature ($\pm=\frac{D}{V}$)	Backward		Rad/μ
700	Path Curvature ($\pm=\frac{D}{V}$)	Backward	Absolute	Rad/μ
701	Path Curvature ($\pm=\frac{D}{V}$)	Backward	Positive	Rad/μ
702	Path Curvature ($\pm=\frac{D}{V}$)	Backward	Negative	Rad/μ