

Published in final edited form as:

*Nat Methods*. 2018 September ; 15(9): 645–646. doi:10.1038/s41592-018-0112-1.

## An open source platform for analyzing and sharing worm behavior data

**Avelino Javer<sup>1,2</sup>, Michael Currie<sup>3</sup>, Chee Wai Lee<sup>3</sup>, Jim Hokanson<sup>3,4</sup>, Kezhi Li<sup>1,2</sup>, Céline N Martineau<sup>5</sup>, Eviatar Yemini<sup>6</sup>, Laura J Grundy<sup>7</sup>, Chris Li<sup>8</sup>, QueeLim Ch'ng<sup>9</sup>, William R Schafer<sup>7</sup>, Ellen AA Nollen<sup>5</sup>, Rex Kerr<sup>10</sup>, and André EX Brown<sup>\*,1,2</sup>**

<sup>1</sup>MRC London Institute of Medical Sciences, London, UK

<sup>2</sup>Institute of Clinical Sciences, Imperial College London, London, UK

<sup>3</sup>OpenWorm Foundation, San Diego, USA

<sup>4</sup>Department of Biomedical Engineering, Duke University, Durham, USA

<sup>5</sup>European Research Institute for the Biology of Ageing, University of Groningen, Groningen, NL

<sup>6</sup>Department of Biological Sciences, Columbia University, New York, USA

<sup>7</sup>MRC Laboratory of Molecular Biology, Cambridge, UK

<sup>8</sup>Department of Biology, City College of the City University of New York, New York, USA

<sup>9</sup>Centre for Developmental Neurobiology, King's College London, London, UK

<sup>10</sup>Calico Life Sciences LLC, South San Francisco, USA

---

Animal behavior is increasingly being recorded in systematic imaging studies that generate large data sets. To maximize the usefulness of these data there is a need for improved resources for analyzing and sharing behavior data that will encourage re-analysis and methodological developments<sup>1</sup>. However, unlike genomic or protein structural data, there are no widely used standards for behavior data. It is therefore desirable to make data available in a relatively raw form so that different investigators can use their own

---

\* andre.brown@imperial.ac.uk.

### Data Availability Statement

Videos, skeleton (WCON) files, and feature files are available with a Creative Commons attribution (CC BY) license through the database page <http://movement.openworm.org/> and Zenodo community page <https://zenodo.org/communities/open-worm-movement-database/>

### Code Availability Statement

Tierpsy Tracker is available as Supplementary Software and at <https://doi.org/10.5281/zenodo.1323782>. Updated versions will be made available at <http://ver228.github.io/tierpsy-tracker/>.

### Author contributions

A.J. wrote Tierpsy Tracker and analyzed data; M.C. wrote WCON viewer, database, and OpenWorm Analysis Toolbox; C.W.L. wrote database, web interface, and WCON viewer; J.H. wrote MATLAB WCON viewer and OpenWorm Analysis Toolbox; K.L. wrote stage alignment code; C.N.M. collected data; E.Y. wrote skeletonization algorithm and stage alignment code; L.J.G. collected data; C.L. contributed strains and planned experiments; Q.C. contributed strains and planned experiments; W.R.S. planned the study; E.A.A.N. contributed strains and planned experiments; R.K. designed WCON and wrote several readers; A.E.X.B. planned the study and wrote the manuscript.

### Competing Financial Interests Statement

The authors declare that they have no competing financial or non-financial interests as defined by Nature Research.

representations and derive their own features. For computational ethology to approach the level of maturity of other areas of bioinformatics, we need to address at least three challenges: storing and accessing video files, defining flexible data formats to facilitate data sharing, and developing software to read, write, browse, and analyze the data. We have generated an open resource to begin addressing these challenges for *C. elegans* behavioral data.

To store video files and the associated feature and metadata, we use a [Zenodo.org](https://zenodo.org) community (an open-access repository for data) that provides durable storage and citability, and that supports contributions from other groups. We have also developed a web interface that enables filtering of the video files based on feature histograms that can return, for example, fast and curved worms in addition to more standard searches for particular strains or genotypes (Fig. 1 and <http://movement.openworm.org/>). The database currently consists of 14,874 single-worm tracking experiments representing 386 genotypes (building on 9,203 experiments and 305 genotypes in a previous publication<sup>2</sup>) and includes data from several larval stages as well as data from ageing experiments consisting of over 2,700 videos of animals tracked daily from the L4 stage to death (see Life Sciences Reporting Summary). Full resolution videos are available in HDF5 containers that include gzip-compressed video frames, timestamps, worm outline and midline, feature data, and experiment metadata. HDF5 files are compatible with multiple languages including MATLAB, R, Python, and C. We have also developed an HDF5 video reader that allows video playback with adjustable speed and zoom (important when reviewing high-resolution, multi-worm tracking data), as well as toggling of worm segmentation over the original video to verify segmentation accuracy during playback.

Secondly, we have defined an interchange format named Worm tracker Commons Object Notation (WCON), to facilitate data sharing and software reuse among groups working on worm behavior. WCON uses the widely supported JSON format to store tracking data as text that is both human and machine readable. It is compatible with single and multi-worm<sup>3</sup> tracking data, at any resolution: from a single point representing worm position over time<sup>4</sup>, to many points representing the high-resolution skeleton of a moving worm<sup>2</sup>. It also supports custom feature additions so that individual labs can store their own specific data sets alongside the existing set of basic worm data. WCON readers are available for Python, MATLAB, Scala, and C. Detailed documentation for the file formats and software is available on the project page (<https://github.com/openworm/tracker-commons>).

Finally, we have complemented the database and file formats with open-source software written in Python for single and multi-worm tracking, feature extraction, review, and analysis (Supplementary Discussion; code and documentation is available as Supplementary Software, and at <https://doi.org/10.5281/zenodo.1323782>, where compiled versions are available in addition).

The suite of tools we have reported makes quantitative behavior (re-)analysis accessible for both experimentalists and computational scientists. It may also serve as a template for similar efforts in other model organism communities.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

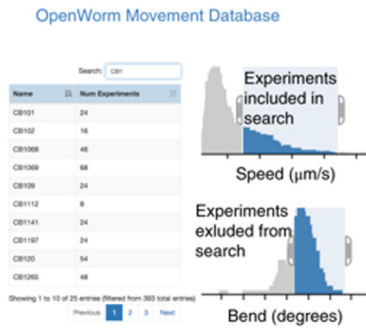
This work was supported by the MRC through grant MC-A658-5TY30 to AEX Brown. Q Ch'ng is supported by an ERC Starting Grant (NeuroAge 242666), Research Councils UK Fellowship, and the University of London Central Research Fund. Some strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

## References

1. Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF. *Nat Neurosci.* 2014; 17:1455–1462. [PubMed: 25349912]
2. Yemini E, Jucikas T, Grundy LJ, Brown AEX, Schafer WR. *Nat Methods.* 2013; 10:877–879. [PubMed: 23852451]
3. Swierczek NA, Giles AC, Rankin CH, Kerr RA. *Nat Methods.* 2011; 8:592–598. [PubMed: 21642964]
4. Ramot D, Johnson BE, Berry TL, Carnell L, Goodman MB. *PLoS ONE.* 2008; 3:e2208. [PubMed: 18493300]

**a** Databases

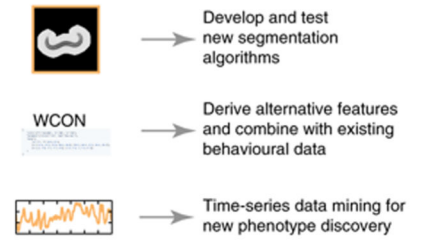
① Search by text of feature



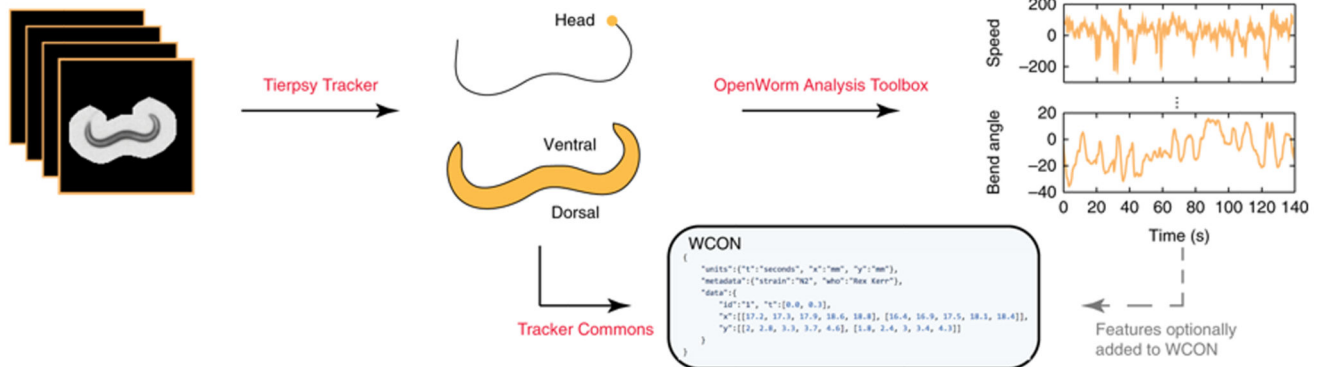
② Download files from repository



③ Process files locally



**b** Software repositories



**Figure 1. Schematic of searchable database and Tierpsy analysis pipeline.**

(a) The OpenWorm Movement Database provides a web interface for searching by genotype, strain, and/or other discrete values, and interactive histograms with sliders to filter results based on feature values. The interface points to data stored on Zenodo. The video and feature data can be further analyzed or combined with data collected using other worm trackers through the Worm tracker Commons Object Notation (WCON), a human and machine readable JSON format. (b) Tierpsy (short for Tierpsychology, the German word for ethology) segments and tracks worms, extracting the outline and skeleton of each animal then determining the head-tail orientation. These data are saved in WCON. The Open Worm Analysis Toolbox is then used to extract behavioral features.