***eLife’s* transparent reporting form**

We encourage authors to provide detailed information *within their submission* to facilitate the interpretation and replication of experiments. Authors can upload supporting documentation to indicate the use of appropriate reporting guidelines for health-related research (see [EQUATOR Network](http://www.equator-network.org/%20)), life science research (see the [BioSharing Information Resource](https://biosharing.org/" \t "_blank)), or the [ARRIVE guidelines](http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1000412) for reporting work involving animal research. Where applicable, authors should refer to any relevant reporting standards documents in this form.

If you have any questions, please consult our Journal Policies and/or contact us: [editorial@elifesciences.org](mailto:editorial@elifesciences.org).

**Sample-size estimation**

* You should state whether an appropriate sample size was computed when the study was being designed
* You should state the statistical method of sample size computation and any required assumptions
* If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

No explicit sample size calculations were made as part of the work described in this manuscript, as the sample size was constrained largely by sample availability and quality, and the cost of sequencing.

**Replicates**

* You should report how often each experiment was performed
* You should include a definition of biological versus technical replication
* The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
* If you encountered any outliers, you should describe how these were handled
* Criteria for exclusion/inclusion of data should be clearly stated
* High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

We chose to focus on sequencing as many worm samples as possible, rather than re-sequencing any worms. The exception is that some worms were sequenced multiple times in an effort to strengthen the reference genome generated based on a limited set of worms. Those worms were selected solely based on the quantity and quality of the DNA extracted. The data availability statement is as follows:

Data are available under the National Center for Biological Information (NCBI) BioProject numbers; PRJNA511012 and SRA submission SUB4949491 for sequencing data, and PRNJA515325 for the genomic assembly.

Links to all genome assemblies are available at: [All](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/Genome_assemblies.tar.gz), [de novo](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version0-genome-assembly.fasta.gz), [semi-de novo (V1)](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version1-genome-assembly.fasta.gz), [V2](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version2-genome-assembly.fasta.gz), [V3](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version3-genome-assembly.fasta.gz), [V4](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version4-genome-assembly.fasta.gz), [V5](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/AL-version5-genome-assembly.fasta.gz) and [mitochondrial](https://s3.amazonaws.com/proj-bip-prod-publicread/his-omics/ALv5/Genome_Assembly/mitochondrial_genomes.tar.gz).

**Statistical reporting**

* Statistical analysis methods should be described and justified
* Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
* For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
* Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

* In most figures, each worm is treated as an individual sample and shown individually.
* All p-values are shown (including those which are not significant in tables 2, S8 and S9).
* Bonferroni corrections are used in Tables 2 and S9. This method was chosen as a cautious choice of multiple comparison correction.

(For large datasets, or papers with a very large number of statistical tests, you may upload a single table file with tests, Ns, etc., with reference to sections in the manuscript.)

**Group allocation**

* Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
* Indicate if masking was used during group allocation, data collection and/or data analysis

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn’t apply to your submission:

No group allocation or randomization was employed. The random forest models used “out of bag” methods, meaning that each sample serves over time as both the training and the testing dataset.

**Additional data files (“source data”)**

* We encourage you to upload relevant additional data files, such as numerical data that are represented as a graph in a figure, or as a summary table
* Where provided, these should be in the most useful format, and they can be uploaded as “Source data” files linked to a main figure or table
* Include model definition files including the full list of parameters used
* Include code used for data analysis (e.g., R, MatLab)
* Avoid stating that data files are “available upon request”

Please indicate the figures or tables for which source data files have been provided:

The datasets consist largely of sequencing data, which is made available publicly. The information about house location, gender, age, etc. is not available publicly as it could be identifying.