

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No code was used to collect data used in this study.

Data analysis

Custom code was generated for geospatial modeling and other analyses. Code was written in Python (version: 3.6.5) and R (version 3.5.1) using Jupyter (version 5.5.0) and R-Markdown (version 1.2.1335). All Google Earth Engine code was written in Javascript. We used the following Python packages: pandas, numpy, sklearn. We used the following R packages: reshape2, corrplot, tidyverse, cowplot, tidyverse, rworldmap, NbClust, factoextra, cluster, vegan, RColorBrewer, lemon, parallel, plotrix, caret, scales. All code is deposited in a GitLab repository linked to this manuscript: https://gitlab.ethz.ch/devinrouth/Crowtherlab_Nematode

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data is deposited in a GitLab repository linked to this manuscript: https://gitlab.ethz.ch/devinrouth/Crowtherlab_Nematode and is attached to the manuscript as Supplementary Table S1. Information from this dataset was used to produce all figures presented in the manuscript.

Supplementary Table S3 lists all 73 covariate layers that were used in geospatial modeling, including their respective sources.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study used nematode abundance and trophic group composition data to characterize the global distribution patterns of soil nematodes across the world's soil. We collected 6,759 soil samples from 1,913 unique locations (1-km ² pixels) worldwide. In each sample we counted number of nematodes and assigned them to trophic groups based on morphology.
Research sample	Each sample consisted of a 15-cm deep soil core, from which soil nematodes were extracted.
Sampling strategy	We collected 6,759 samples from 1,913 unique locations (1-km ² pixels) worldwide. Samples were collected from published or personal datasets. No sample size calculation was performed, the sample size was deemed to have sufficient spatial coverage compared to that used in previous global soil biogeographic studies. The representativeness of the sample set was assessed using univariate and multivariate analysis across the 73 covariate layers included in the modeling.
Data collection	Nematode abundance data was collated from published and personal datasets.
Timing and spatial scale	Samples were collected from all continents and all biogeographic regions on Earth. The bulk of samples was taken between 2002 and 2017, full range of sampling dates was from 1963 to 2017.
Data exclusions	After collection, all samples lacking spatial meta-information (i.e., lat/long coordinates) were excluded from the dataset. After aggregation into 1-km ² pixels, pixels classified as water bodies or snow/ice covered were excluded. Any points that fell off the pixel grid due to locational coordinate errors were excluded. Antarctic samples were excluded from geospatial modeling due to limited coverage of the covariate layers. For inter-trophic group correlation calculations, only samples with full observations (i.e., reported counts for each trophic group) were included.
Reproducibility	Each model's strength was tested by K-fold cross validation (using k=10). As such, we assessed the reproducibility of the model on 10 unique training datasets, providing high confidence (R ² 0.86, mean k-fold R ² 0.43) in the final model results. Spatially-explicit coefficient of variation-values, as a measure of prediction accuracy, were calculated for each pixel by calculating the standard deviation and mean values across each of the 10 folds. All random seeds were specified manually and written in the code. To create statistically valid per-pixel confidence intervals, we performed a stratified bootstrapping procedure with the "total number" collection of nematode point data. The stratification category was the sampled biomes of each point feature (to avoid biases), and the number of bootstrap iterations was 100, in each of the bootstrap iterations classifying the composite raster data 100 times. This protocol thus allowed us to generate per-pixel 95% confidence intervals.
Randomization	Samples were included in each trophic group analyses per their inclusion in trophic level information.
Blinding	No treatments groups were assigned in the study, as we performed an observational and exploratory geospatial mapping project. Hence, we deem there is no need for blinding.
Did the study involve field work?	☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

N/A

Wild animals

This study involved field-sampled microscopic nematodes.

Field-collected samples

This study involved field-sampled microscopic nematodes that do not require special institutional review for ethics.